



QUALITATIVE COMPARATIVE ANALYSIS IN CRITICAL PERSPECTIVE

Samuel R. Lucas*

Alisa Szatrowski*

Abstract

Qualitative comparative analysis (QCA) appears to offer a systematic means for case-oriented analysis. The method not only offers to provide a standardized procedure for qualitative research but also serves, to some, as an instantiation of deterministic methods. Others, however, contest QCA because of its deterministic lineage. Multiple other issues surrounding QCA, such as its response to measurement error and its ability to ascertain asymmetric causality, are also matters of interest. Existing research has demonstrated the use of QCA on real data, but such data do not allow one to establish the method's efficacy, because the true causes of real social phenomena are always contestable. In response, the authors analyze several simulated data sets for which true causal processes are known. They find that QCA finds the correct causal story only 3 times across 70 different solutions, and even these rare successes, on closer examination, actually reveal additional fundamental problems with the method. Further epistemological analyses of the results find key problems with QCA's stated epistemology, and results indicate that QCA fails even when its stated epistemological claims are ontologically accurate. Thus, the authors conclude that analysts should reject both QCA

*University of California, Berkeley, USA

Corresponding Author:

Samuel R. Lucas, University of California, Berkeley, Sociology Department, 410 Barrows Hall
#1980, Berkeley, CA 94720-1980, USA

Email: lucas@berkeley.edu

and its epistemological justifications in favor of existing effective methods and epistemologies for qualitative research.

Keywords

QCA, identification, asymmetric causality, case-oriented research, determinism, epistemology

1. INTRODUCTION

Skocpol's (1979) landmark *States and Social Revolutions* reinvigorated interest in the study of macro-level phenomena via the comparative assessment of historical cases. That comparative historical researchers often access detailed, complex, nonstandardized information, such that statistical analysis is often impossible, has motivated some analysts to develop standardized, nonstatistical analysis techniques. With crisp-set and fuzzy-set qualitative comparative analysis (QCA), Ragin (e.g., 2005) offers such a systematic tool.

According to Ragin (2008a), QCA purports to allow case-oriented research using more cases than usual. QCA analysts identify causal recipes—sets of characteristics—and may discern asymmetric causes. *Sociological Methods and Research*, *Studies in Comparative International Development*, and *Sociological Methods and Theory* have published symposia on QCA, spurring growing diffusion and acceptance of the method. In the following pages, however, we sound a strongly discordant note.

The tone of that discordant note need be clear: We are not critical of comparative historical research specifically or of qualitative research in general. We find many such analyses incredibly rich and illuminating. We ask not whether comparative historical or other qualitative research is illuminating but, instead, whether transforming rich qualitative data into a form amenable to a Boolean or fuzzy-logic engine and analyzing those data using such an engine is illuminating, and we use the most widely used such engine, QCA, to ask the question.¹

Further, we do not assess the value of logic methods for logical analysis. Analysts often identify or construct logical relations (e.g., if $A = B$, then $C = 1$; if $A \neq B$, then $C = 0$). Boolean or fuzzy logic can aid in analyzing such cases. The world of logic and the world of empirical experience, however, are not the same. We ask only whether these methods are useful for empirical analysis.

Through several simulations, we find that the answer to our question is no: The engine of QCA runs, but it takes one to erroneous conclusions

time and time again. That is, QCA fails to find correct causal recipes, fails to replicate causal recipes across data sets that differ only owing to chance, identifies causal patterns in noncausal data, does not find the correct causal patterns in deterministic data, finds interactions even when they are absent, fails to find the correct interactions when interactions are present, selects the wrong direction of association, and finds asymmetric causation when the known causal structure is symmetric. All these disappointments occur despite our providing the correct causal variables in every simulation save one (a test using noncausal data), reducing the chance of erroneous findings. Given these results, we conclude that QCA is a wholly ineffective research method, providing a fatal distraction that, far from realizing the promise of qualitative and comparative historical research, sabotages such analyses.²

We further find that what many analysts understand QCA to do is very different from what the method's inner workings actually do. This is unfortunate, for QCA draws evaluative interest not only because it appears to offer a way to systematically analyze qualitative data but also because it could contribute to the smoldering debate around fundamental epistemological issues. In such debates, QCA users and critics take what the method is understood to do as given and focus on whether what the method is understood to do is advisable. However, because the method does not operate as claimed, its contribution to epistemological debate has actually been to obfuscate matters. As we proceed, we will trace the epistemological implications of our findings.

We begin by briefly describing QCA and then outlining debates about the method. Next we justify our use of simulations, an approach QCA analysts have rejected. After an introductory illustrative real-world analysis, we present six simulations, noting the implications after each (five supplementary analyses and data [or code for constructing data] for the 11 studies are available online at <http://sm.sagepub.com/supplemental>). In the penultimate section, we address a foundational issue that emerges from our analyses, and we then close with summary conclusions.

2. QUALITATIVE COMPARATIVE ANALYSIS

2.1. *Using QCA*

We briefly describe QCA below. (For didactic explications see, variously, Berg-Schlosser et al. 2009; Ragin 2000, 2008a, 2008b, 2009; Rihoux and Grimm 2006; Rihoux and Ragin 2009). QCA researchers

seek to determine the causal recipes that produce the occurrence (e.g., $Y = 1$) and nonoccurrence (e.g., $Y = 0$) of an outcome; indeed, they investigate occurrence and nonoccurrence separately to assess asymmetric causality, which is understood to mean that the opposite of the cause of $Y = 1$ (or $Y = \text{high}$) is not necessarily a cause of $Y = 0$ (or $Y = \text{low}$). Causal recipes are composed of exact values of selected variables; for example, if dichotomous variables A , B , and C might cause dichotomous $Y = 1$, one possible causal recipe is $A = 1$, $B = 0$, $C = 1$, while $A = 1$, $B = 1$, $C = (0 \text{ or } 1)$ is another. Different causal recipes may cause the same outcome, reflecting the presence of equifinality.

Analysts may use crisp-set QCA (csQCA) on binary variables that indicate membership in a crisp set with sharp boundaries. However, fuzzy-set QCA (fsQCA) can use continuous variables recoded into ordered categories of fuzzy-set membership. Whereas crisp sets do not allow degrees of membership, fuzzy sets do. For example, a fuzzy-set measure might classify nations as full members in the set of nuclear powers, mostly in that set, about equally inside and outside that set, and so on down to fully outside that set. In contrast, in crisp-set terms, nations either have nuclear weapons or do not.

2.2. Analyzing Data Using QCA

Although csQCA and fsQCA are similar, one key difference between them is that fsQCA requires analysts to calibrate their measures. Ragin (2008a:88–97) proposes two calibration strategies, with each mapping a variable onto the logistic distribution. The first strategy anchors key substantive points at theoretically important values and then arrays the remaining values along that continuum, after which an analyst uses the logistic distribution to transform the data into a scale that asymptotically approaches zero and one. A second strategy transforms the variable using the logistic distribution directly, without theoretically based anchors. Because Ragin (2008a) highlights theory, the first method is preferred. Either way, some values short of the asymptotes are selected to mark complete membership and nonmembership in the set.³

Once calibration is accomplished, csQCA and fsQCA analysts may follow the same procedures. Analysts first construct a truth table, a table containing a row for each permutation of the variables (e.g., one row for $X_1 = 0$, $X_2 = 0$, $X_3 = 0$, one row for $X_1 = 1$, $X_2 = 0$, $X_3 = 0$, and so on; an fsQCA truth table row reflects a corner in multidimensional

space [Ragin 2008a:126–33]). Each row also indicates the frequency of cases matching the permutation, and the proportion of cases in the row for which the outcome is observed (e.g., $\Pr[Y = 1]$).

Analysts next remove from the table all configurations that fall below a chosen *frequency-of-cases* threshold and then code as causal for the outcome all configurations that exceed a chosen *consistency threshold*, defined as the proportion of cases for which the outcome (e.g., $Y = 1$) is observed. Analysts are advised to use a consistency threshold of .75 or higher (Ragin 2008a:143–44). On this basis, QCA produces a complex solution.

Some combinations of variable values may lack cases, meaning that some configurations that could provide useful analytic leverage may not be observed. For example, there may be no Marxist, Buddhist, European Union states. A QCA analyst may assign to nonexistent (i.e., counterfactual) configurations the outcome the analyst believes the configuration would produce if the configuration were to exist. Once the analyst makes these assignments, a more parsimonious intermediate solution based on invoking those assignments may be produced. A still more parsimonious solution is produced by QCA's mechanical imposition of additional simplifying assignments to any remaining nonexistent configurations. The complex solution does not depend on the assignment of such counterfactual configurations. Analysts are encouraged to highlight the type of solution most appropriate for their research (Schneider and Wagemann 2010:408). Because of length, in what follows we usually highlight the complex solution, for it reflects fewer *a priori* researcher assumptions. Nevertheless, we often present multiple solutions to help readers evaluate our results, and sometimes we discuss more than just the complex solution.

QCA also reports the level of consistency and coverage of the solution and of each causal recipe. In crisp-set analyses, consistency reflects the proportion of cases for which the outcome occurs and coverage reflects the percentage of all outcome occurrences associated with the causal recipe(s). In fuzzy-set analyses, the mathematics differs, but consistency and coverage are conceptually analogous in crisp-set and fuzzy-set analyses.⁴

Prior to fsQCA, QCA analysts needed to transform all data into dichotomies. Ragin (2008a:141) now recommends fsQCA unless the data possess a strong nominal character.

Table 1. Selected Mill's Methods

Name	Definition	Example
Method of agreement	"If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon" (Mill [1843] 1906:255).	If two nations have interethnic violence, and the only circumstance they share is the same language, then this circumstance is the cause or the effect of interethnic violence.
Method of difference	"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance in common save one, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or the cause, or an indispensable part of the cause, of the phenomenon" (Mill [1843] 1906:256).	If two nations are under study, and they differ in that revolution occurs in one instance but not in the other, the only candidates for causing the revolution are those factors that differ between the two nations.

2.3. Stated Epistemological Bases of QCA

QCA is built on the epistemological base of John Stuart Mill's methods (Ragin 1987); two key such methods are conveyed in Table 1.

Describing QCA as "set theoretic," conceiving of variables as indicators of set membership and of relations between (groups of) variables as relations between sets, QCA analysts attempt to discern sufficient causes, necessary causes, and combinations thereof (see Table 2), where necessary causes are in all causal recipes but sufficient causes are not. Thus, QCA is seen as revealing explicit causal connections between sets of conditions and outcomes.⁵

Another key claim, described as a fundamental difference between QCA and quantitative correlation analyses (e.g., Ragin 2008a:102, 139), is that QCA allows for asymmetric relationships between causal conditions and outcomes. Illustratively, statistical researchers are said to assess the causal role of a dichotomous X on a dichotomous Y by using the odds ratio of the two-by-two cross-tabulation table. The odds ratio compares the distributions of $Y|X = 1$ and $Y|X = 0$. If they differ, then X

Table 2. Four Forms of Causation

Name	Definition	Example
Necessary	A cause that is always present in the case of the outcome. However, sometimes the cause will be present but the outcome will not be observed.	If external war is a necessary cause of revolution, then all cases of revolution will occur in states engaged in external war. However, some states engaged in external war may not experience revolutions.
Sufficient	A cause that always produces the outcome, though other causes may also produce the outcome. Sometimes the outcome will be present but the cause will be absent.	If state repression is a sufficient cause of austerity protest, then the outcome (austerity protest) must arise in every instance in which the sufficient cause (state repression) occurs. Still, it would be possible for an austerity protest to occur even without state repression
INUS	A cause X that is an insufficient but necessary part of condition Z , where Z is a sufficient cause of Y .	If failed harvest is a necessary part of sufficient condition Z , composed of failed harvest and civil war, and Z is a sufficient cause of revolution, then all cases with failed harvests and civil war will cause revolution, revolution will also be caused without sufficient condition Z , and civil war without failed harvests but some other factor may also cause revolution, and revolution may also occur in the absence of civil war.
SUIN	A cause Q that is a sufficient but unnecessary part of P , where P is a necessary cause of Y .	If failed harvest is a sufficient but unnecessary part of condition P , composed of failed harvest and civil war, and the condition is a necessary cause of revolution, then all revolutions will occur in states with failed harvests regardless of the existence or nonexistence of civil war, some states will have revolutions despite successful harvests such that failed harvest/civil war is an unnecessary case of revolution.

and Y are associated and a causal relationship between X and Y is possible. In studying these distributions, statistical analysts consider all cells of the table.

In contrast, Ragin (2008a) contends that some table cells are irrelevant to whether X causes Y , arguing that analysts interested in causal conditions shared by an outcome will consider cells in which $Y = 1$ (cells $[0,1]$ and $[1,1]$) and that analysts interested in outcomes produced for cases sharing the same causal condition would consider cells in which $X = 1$ (cells $[1,0]$ and $[1,1]$). The cell signifying absence of both cause and outcome ($X = 0, Y = 0$) “is not directly relevant to the assessment of either of the two types of explicit connections” (Ragin 2008a:22). In this way, QCA analysts find asymmetric causal relations other analysts cannot.

For proponents, QCA is a set-theoretic method that embraces asymmetric relationships and detects causal links. The contention is that new, perhaps deeper insight can be gained by using such methods (Fujita 2009). Indeed, Ragin (2008a:190–212) contends that QCA can be applied to very large data sets and, addressing Liao (2001), illustrates the utility of QCA by reanalyzing more than 12,000 cases from the 1979 National Longitudinal Survey of Youth (NLSY) to reassess Herrnstein and Murray’s (1994) *The Bell Curve*.

3. EPISTEMOLOGICAL DEBATE AROUND QCA

3.1. *Determinism and Stochasticism*

Much of the debate about the utility of QCA pivots around one key epistemological disagreement: Are deterministic analyses useful or not?⁶ This central issue is debated directly and is evident in two additional issues: (1) whether QCA uses Mill’s methods and to what effect and (2) whether QCA is undone by measurement error.

Mahoney (2008) embraces both determinism and QCA as a deterministic technique, arguing that case-oriented researchers use several variations on necessary and sufficient causation (see again Table 2). He then writes that adopting any of these types of causality is

to assume that the world of human beings is deterministic. This assumption does not of course mean that a given researcher will be successful at identifying the causes of any outcome. Nor does this assumption mean that the researcher will achieve valid measurement or carry out good research. The

assumption of an ontologically deterministic world in no way implies that researchers will successfully analyze causal processes in this world. But it does mean that randomness and chance appear only because of limitations in theories, models, measurement, and data.

The only alternative to ontological determinism is to assume that, at least in part, “things just happen”; that is, to assume that truly stochastic factors—whatever those may be—randomly produce outcomes. (p. 420, endnotes and citations suppressed)

Mahoney (2000) identifies necessary and sufficient causation as deterministic, and he indicates (Mahoney 2008) that QCA, as a means to study necessary and sufficient causes, is deterministic.

An additional claim is that case-oriented researchers are attempting to understand a specific case or set of cases (Mahoney and Goertz 2006). In that view, the French Revolution—or one roll of the dice—is a deterministic phenomenon in that the probability of the outcome we observe is 1, and the probability of all other outcomes is 0 (Mahoney 2008:415–16). Seen in this way, case-oriented researchers seek a comprehensive explanation, that is, an explanation that completely accounts for what is observed. This aim is contrasted with efforts to estimate the effects of variables in populations (Mahoney 2008:420); the claim is that in such studies, multiple cases are deployed to estimate population average effects.

Mahoney embraced determinism, but some analysts, seeing determinism as undesirable for many reasons, prefer stochasticism. First, determinism implies that analysts could offer a nontrivial explanation that would explain every case perfectly (Lieberson 1994, 2004).⁷ Second, an effort to establish full explanations easily leads analysts to capitalize on chance data patterns, that is, to overfit the data (Lieberson 2004). Findings thus produced are unlikely to be replicated. Third, if measurement error exists, analysts will misidentify causes (Goldthorpe 2000; Lieberson 1991, 1994). Thus, the only way to use QCA is to assume perfect measurement.

Ragin (2000, 2008a), attempting to strike a middle position, contends that fsQCA provides a stochastic tool and noted that “common data and evidence problems (error, randomness) provide a very strong motivation to employ analytic techniques that make some use of probability theory, especially techniques that address the problem of drawing inferences from imperfect evidence” (Ragin 2000:109). However, Ragin (2000:15) also asserts the “centrality of the analysis of . . . necessary and sufficient

causation to social scientific inquiry,” a contention that seems to require a deterministic epistemology. Ragin (2000, 2008a) attempts to resolve the two positions by referencing the concepts of almost necessary cause and almost sufficient cause.

Clearly, different analysts have different postures on determinism and stochasticism. Thus, when Lieberman (2004) states that QCA is deterministic, he regards the critique as conclusive, but Mahoney, unfazed by the criticism because he embraces QCA as a deterministic method that matches his ontological commitments, sees QCA as a breakthrough.

Two other issues follow from the division between determinism and stochasticism. QCA proponents view QCA as a proper implementation or extension of Mill’s methods (e.g., Mahoney 2008; Ragin 2008a; Wagemann and Schneider 2010). Others see QCA as flawed to the extent it implements Mill’s methods in social analysis (e.g., Lieberman 1991).

Critics have claimed that with no allowance for measurement error, QCA may easily produce incorrect results (e.g., Goldthorpe 2000; Lieberman 1991). Ragin (2000), however, introduces probabilistic criteria, possibly muting this threat.

3.2. Other Central Aspects of Debate

The above discussion references central points of contention, but others exist as well. Were it not for the mega-issues noted above, these additional points of contention would themselves be major foci for analytic attention. Hence, these smaller (but still large) issues demand mention as well.

QCA analysts argue that to assess asymmetric causation, one must separately analyze $Y = 1$ and $Y = 0$ (or $Y = \text{high}$ and $Y = \text{low}$) (De Meur and Rihoux 2002) and ignore some combinations of the putative causal variable and the outcome (Ragin 2008a:20–23). These counsels run counter to common claims of statistical analysts (e.g., Grofman and Schneider 2009:669).

Another argument is that QCA searches for invariant causal configurations, which creates doubt concerning the causal status of interaction effects for, by definition, interaction effects imply that causal configurations are not invariant (Lieberman 1991, 1994). In response, Savolainen (1994) contends that interactions are incredibly flexible, and the only way to sustain Lieberman’s critique is to deny that complexity. More

pointedly, Vaisey (2007) claims that there is a major difference between interactions and complex conjunctural arrangements, and QCA allows assessment of the latter.

Another argument is that QCA is no advance over statistical modeling because astute statistical modelers already follow QCA advice to attend to context, base measurement on theory, and more (Achen 2005; Seawright 2005). QCA supporters respond by contending that QCA is meant to supplement, not supplant, statistical modeling (Ragin and Rihoux 2004).

Another argument is that QCA requires restrictive or contradictory assumptions for omitted variables and a functional form assumption that is as necessary as the functional form assumptions statistical analysts must make (Seawright 2005). Yet QCA analysts point to calibration in fsQCA as a response to the functional form issue (Ragin 2008a).

Another argument is that QCA abandons useful features of statistics such as standard errors (Achen 2005:30; Seawright 2005:14), prompting Seawright to suggest that the insights of QCA—most notably interest in complex conjunctures and nonlinear functional forms—be integrated into other methodological frameworks, including both statistical analysis and other methods of case-study research. Yet Ragin (e.g., 2000:107–18) shows that one may use statistical testing in QCA, undercutting this claim

In what follows we address these contentious issues.

4. METHODS

QCA analysts urge the use of QCA on real-world data. Confining use of QCA to real-world data is unfortunate. Because QCA purports to instantiate a deterministic epistemology, study of QCA's behavior under controlled conditions could aid efforts to adjudicate seriously contentious epistemological claims. Yet QCA has not served this purpose for two reasons.

First, reanalyzing data formerly analyzed by some other method cannot, by itself, reveal the validity of QCA, because any outcome may be dismissed. If reanalysis fails to replicate earlier results, critics can contend that this shows that QCA is unsuccessful, while proponents can claim that it reveals the power of QCA to produce new knowledge. However, if the reanalysis replicates earlier results, proponents can contend that this proves that QCA is solid, while critics can contend that

this indicates that QCA is redundant, merely rediscovering facts we can already excavate with existing methods. Thus, using QCA methods on real data (e.g., Fujita 2009; Grofman and Schneider 2009) ultimately cannot validate the method.

Second, any method can produce incorrect findings; for example, one may have in hand one of the few exceedingly atypical probability samples. When the data are drawn from the real world, findings cannot be independently validated. Thus, one can always contest the utility of any proposed method by claiming that its findings on real-world data are wrong.

One response is to assess whether QCA identifies the causal story on data for which the causal story is already known. The only such data are simulated data.

Lieberson (2004) proposed three such simulation-based tests for QCA. However, Ragin and Rihoux (2004) reject simulations, claiming that QCA is for researchers intrinsically interested in the cases they study. Others seem to agree, claiming that failure to return to the specific cases is to misuse QCA (Schneider and Wagemann 2010). When cases are simulated, there is precious little to turn to outside of the simulated variables, for even omitted variables will have been fabricated. Thus, the claim suggests that simulations are useless for evaluating QCA.

Yet some QCA analyses do not return to the cases. For example, in fsQCA analysis of NLSY data, Ragin (2006, 2008a) does not interpret fsQCA results by returning to the real cases or even to the hundreds of variables in the NLSY that were omitted from fsQCA analysis. In Dy et al.'s (2005) QCA study of 26 critical hospital pathways to assess the determinants of effectiveness, the researchers also do not return to the cases to aid interpretation. Several such examples exist (e.g., Glaesser et al. 2009), indicating that QCA does not require a return to cases for interpretation and thus implying that simulations focused only on the truth table are appropriate to test key aspects of QCA.

Resistance to simulations seems to deny that given J simulated cases such that

$$Y_j = f(\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \varepsilon_j), \varepsilon_j \sim \eta(u, \sigma^2), \quad (1)$$

with f a linear, nonlinear, or nonparametric function, each β a fixed value or vector, ε_j set as random with distribution η or degenerate at zero if deterministic causality is specified, with mean μ and variance

σ^2 , and each X_{jk} with nonzero variance allocated values however desired, then

$$E(Y_j) = f(b_0 + b_1X_{1j} + b_2X_{2j} + b_3X_{3j}) \quad (2)$$

is the completely correct causal specification for the determination of Y , where ε_j is unobserved by the stochasticist or zero for the determinist.⁸ Any method that cannot identify the correct causal story under these conditions—where perhaps 100 percent of the pertinent information is available—is unlikely to succeed in real-world research. Hence, simulations are of potential value for evaluating the effectiveness of QCA. Still, because of QCA proponents' concerns, we located a high-quality real-world data set for which diverse analysts agree about cause. We will therefore demonstrate the concerns and questions that motivate our simulations by first analyzing these data.

Another knotty issue is that Monte Carlo simulations rely on probability logic. QCA analysts' use of nonprobability samples (e.g., Hollingsworth et al. 1996) suggests that they may reject probability theory and thus find results produced thereby unpersuasive. Setting aside that nonprobability samples are nearly useless unless one is interested only in the cases studied (Lucas 2014a), and other criteria are met (Lucas 2014b), we highlight results based on methods that do not use probability theory.

Thus, in the following pages, we proceed as follows. After a brief analysis of real-world data, we present our simulations. For each simulation, we relate the data, their method of construction, the results, and then their implications. We note that QCA advice, coupled with our interest in full consideration, greatly multiplies the analyses one might desire. For example, QCA advises separate analyses of $Y = 1$ (or high) and $Y = 0$ (or low). As another example, although csQCA and fsQCA appear similar, we are predisposed to check every result across methods. These two features quadruple the number of analyses, threatening to greatly lengthen the article and blur our focus. We have two responses to this challenge. First, we alter our methods as we learn which concerns can be set aside, tightening and shortening the analysis. Second, we focus our work by designing each simulation to study issues of enduring methodological interest, common real-life research challenges, and/or specific claims proponents offer for QCA, using the advantage that data with known causal relations afford. The six simulation studies here

investigate fundamental aspects of research and of the QCA debate. But we first turn to an analysis of real-world data.

5. A QUALITATIVE COMPARATIVE REANALYSIS OF SHUTTLE LAUNCH DATA: AN ORIENTING ILLUSTRATIVE INVESTIGATION

Vaughan (1996) studied the process by which the decision to launch STS-51-L—the *Challenger* shuttle—on January 28, 1986, was made, concluding that the distal cause was a NASA culture that normalized risk. Engineers with Morton Thiokol, a subcontractor, warned against launch given unusually low forecast temperatures. Yet they lacked data sufficient to convince NASA managers to delay in a context also stressed by politicians' demands for routinized launch schedules (Vaughan 1996:210). With NASA resisting, Morton Thiokol managers overruled their own engineers and supported launch. Alas, as engineers feared, cold weather hardened the O-rings, the O-rings thus failed to seal, and hot gases escaped and breached the fuel tank, causing the catastrophic failure of STS-51-L (Robison et al. 2002; Rogers et al. 1986:72; Tufte 1997; Vaughan 1996).

We reanalyze Rogers et al.'s (1986:129–32) data, which engineers had sought but not received before STS-51-L (Robison et al. 2002). Variables indicate whether O-ring erosion occurred, whether blow-by occurred, field and nozzle joint pressures, and joint temperatures for 23 of the 24 pre-STS-51-L launches (one mission's materials were lost at sea). Erosion and blow-by signal that the design is not working as expected, and thus the mission may have been in serious danger (Feynman 1986:F1). We code anomalies noted only in footnotes as nonerosion/non-blow-by, omit tests because operational launches entail real launch processes and conditions (Robison et al. 2002:77), and omit STS-51-L to replicate investigators' analyses. We consider pressure because upon finding erosion in early shuttle flights, pressures were raised even as doing so increased the chance that dangerous blow holes would form in insulating putty (Rogers et al. 1986:133–34). Pressure was studied as a possible cause of erosion (Robison et al. 2002) but was eliminated as a cause of the disaster (Rogers et al. 1986:40). Indeed, even those who disagree about whom to blame agree both on what caused the disaster and that pre-STS-51-L data, properly analyzed, reveal the risk of low-temperature launch independent of other factors

(Robison et al. 2002; Tufte 1997). Thus, pre-STS-51-L launches seem to offer a rare real-world data set of known causes, making them potentially useful for assessing QCA.

Potential explanatory factors are continuous, so we use fsQCA. We calibrate pressure by setting 50 as the lower bound, 125 as the midpoint, and 200 as the upper bound. Because shuttle launch criteria required that temperature fall between 31°F and 99°F (Vaughan 1996:309), we set joint temperature bounds at 32°F and 98°F, with a crossover at 65°F because joint temperature is related to ambient temperature. We use a one-case frequency threshold and a consistency threshold of .75. We present results for erosion and nonerosion only, because fsQCA could identify no cause for blow-by.

We investigate the existence ($Y = 1$) and nonexistence ($Y = 0$) of erosion separately, as QCA literature advises (see Table 3). The fsQCA complex solution for erosion is a three-way interaction of low temperature, high field pressure, and high nozzle pressure and for nonerosion is the opposite, a three-way interaction of high temperature, low field pressure, and low nozzle pressure. The parsimonious solution for erosion is the interaction of high nozzle pressure and low temperature and for nonerosion is low field pressure alone. Thus, QCA finds that temperature has no independent effect, posing a danger only when low while field and nozzle pressure are high.

Many previous analyses of erosion were qualitative studies rather than statistical analyses (Vaughan 1996:354–55, citing Biosjoly). Previous analyses identified low temperature as a cause of danger independent of field and nozzle pressure. Yet QCA identifies low temperature as an issue only in concert with high field and/or nozzle pressure. The difference between QCA findings and other research is consequential. If QCA results are right and the other research is wrong, and cold is not a problem when pressures are low, then NASA could lower field and/or nozzle pressures to address the danger of cold-weather launch. If QCA is wrong and the other research is right, cold poses an overriding danger regardless of field and nozzle pressures, and thus cold-weather launches should be postponed, as Morton Thiokol engineers argued.

If we believe the previous work, then QCA clearly fails in real-world testing. However, one might argue that the earlier research used methods that cannot discern at least some aspects of the true cause, and thus QCA reveals new knowledge. Either way, the real-world result casts our methodological questions of the utility of QCA in bold and potentially

Table 3. Results of Shuttle Launch O-ring Erosion Reanalysis via QCA (*n* = 23)

Solution	Erosion		Nonerosion	
	Causal Recipe	Coverage, Consistency	Causal Recipe	Coverage, Consistency
Complex	Low temperature × high field pressure × high nozzle pressure	.359, .759	High temperature × low field pressure × low nozzle pressure	.396, .784
Parsimonious	Low temperature × high nozzle pressure	.359, .759	Low field pressure	.583, .805

even fatal relief: What leads QCA to reach conclusions that oppose Boisjoly's and other engineers' preflight concerns, the Presidential Commission report, Vaughan's (1996) qualitative case study, and Tufte's (1997) graphical reanalysis? Because simulations can aid in excavating the implemented assumptions and operations of methods, and thereby facilitate assessment of them, it is to the simulations that we now turn.

6. STUDY 1: EQUIFINALITY, NONCAUSALITY, AND DETERMINISTIC CAUSALITY

For study 1, we produced 40 cases with four X variables randomly assigned values of 0 or 1. We constructed $Y = 1$ if $X_1 = X_2 = 1$ or if $X_3 = X_4 = 1$, reflecting equifinality. We also constructed Z_1 noncausal for Y . No error term was used in producing Y ; thus, Y is deterministically caused. We use the full 40 cases of the population. The data thus match Mahoney's vision of causality.

We conducted two csQCA analyses. One uses only the X variables, mimicking a situation in which the analyst knows all determinants of Y ; the other adds a Z variable, mimicking the common situation in which the analyst erroneously includes a noncausal variable in the explanation. A solid method should find that Z is not causally connected to Y .

We used csQCA, the truth-table approach, a consistency threshold of .8, and a one-case frequency threshold. These are program defaults, so they match advice for the use of QCA.

6.1. Study 1 Results

We describe Table 4 in detail, for its features are common to many of the later tables. The first column identifies the complexity of the solution. Next, the "R" column indicates how many truth-table rows were read for the analysis (Ragin 2008b:68). In the next column, each line in the cell is a causal recipe. The variable name preceded by \sim indicates that the absence of the condition is part of the causal recipe; the variable name without \sim indicates that the presence of the condition is part of the causal recipe, while an asterisk indicates "and" and thus joins two causal factors. The table's final column contains the coverage and consistency statistics.

Table 4. Study 1, QCA Analysis with Equifinal Causes ($n = 40$)

Solution	Outcome $Y = 1$			Outcome $Y = 0$		
	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
Complex	14	$x3*x4$ $x1*x2*\sim x4$.8750, 1.0000	14	$\sim x2*\sim x4$ $\sim x1*\sim x4$	1.0000, 1.0000
Intermediate	6	$x3*x4$ $x1*x2*\sim x4$.8750, 1.0000	12	$\sim x1*\sim x2*\sim x3$ $\sim x4*\sim x1$ $\sim x4*\sim x2$ $\sim x3*\sim x2*\sim x1$	1.0000, 1.0000
Parsimonious	14	$x3*x4$ $x1*x2*\sim x4$.8750, 1.0000	14	$\sim x2*\sim x4$ $\sim x1*\sim x4$ $\sim x2*\sim x3$ $\sim x1*\sim x3$	1.0000, 1.0000

Considering the analysis of $Y = 1$, one causal recipe from the complex solution is “ X_3 and X_4 ”; given the coding of the variables, this can be translated as $X_3 = X_4 = 1$. Another causal recipe is “ X_1 and X_2 and not X_4 ,” that is, $X_1 = X_2 = 1$ while $X_4 = 0$. None of the factors are necessary causes of $Y = 1$ or $Y = 0$, for the same value of a given X does not appear in all causal recipes. The complex solution is identical to the intermediate and parsimonious solutions.

Although csQCA identified one of the two causal recipes ($X_3 = X_4 = 1$) correctly, it incorrectly identified the other, claiming that the value of X_4 was relevant for whether $X_1 = X_2 = 1$ is causal for Y . We know that this is incorrect, by construction. This failure is especially concerning, because this is exactly the case for which QCA is advertised: equifinality, deterministic data, and analysts’ knowing the set of causal factors (on the basis of theory and/or extensive case-specific study).

Following advised procedure, we then analyze $Y = 0$. Given how we constructed $Y = 1$, the fundamental causal recipe of $Y = 0$ is $X_1 \times X_2 = 0$ and $X_3 \times X_4 = 0$, which occurs if $X_1 = 0$ and X_3 or $X_4 = 0$ or if $X_2 = 0$ and X_3 or $X_4 = 0$. Results for the $Y = 0$ analysis are not clearly opposite those obtained for $Y = 1$. Thus, QCA appears to find asymmetry.

Indeed, the asymmetry is intriguing, for although the complex and intermediate solutions to $Y = 0$ and $Y = 1$ are wrong, the parsimonious solution for $Y = 1$ is wrong but for $Y = 0$ is correct. The half-right, half-wrong pattern of parsimonious results raises serious questions of QCA.

The analysis in Table 4 uses all causal factors and only causal factors. Analysts sometimes mistakenly omit a causal factor or include a non-causal factor. The costs of omitting a cause are well known for many other methods (e.g., Goldberger 1991:189–91), and QCA should not be held to the untenable standard of producing correct causal conclusions in such a case.

However, if analysts cannot use methods to reveal that factors they believe matter do not matter, empirical analysis of observational data is of little value. Statistical and qualitative analysts routinely appear able to eliminate some factors as causal (e.g., Hauser and Anderson 1991; Vaughan 1996). It is fair to require QCA to succeed at this task as well. To proceed, we added a dichotomous variable, Z_1 , to the analysis. Z_1 , the sum of a random component, X_1 , X_2 , X_3 , and X_4 , is dichotomized at .49 to produce Z_1 ; Z_1 is correlated at .198, .156, .383, .088, and .245 with X_1 to X_4 and Y , respectively. Being constructed later than Y and having the

lowest correlation with Y , Z_1 has no causal role in Y and is not a viable candidate for causal priority.

The short answer from Table 5 is that QCA does not eliminate the noncausal factor from the causal recipes. In fact, not only does QCA fail to eliminate the noncausal factor, it identifies eight four-way interactions as causal recipes in the complex solution for $Y = 1$, five of which involve Z_1 , while half of the four complex solutions for $Y = 0$ also involve Z_1 . These results are produced even though X_1 through X_4 deterministically cause Y . Far from identifying Z_1 as noncausal, QCA deeply embeds the noncausal Z_1 into the causal story.

Note, however, that the parsimonious solution for $Y = 0$ is correct, whereas its counterpart parsimonious solution for $Y = 1$ remains incorrect. This further deepens the questions about QCA.

Faced with these results, one might wonder whether the data actually follow the pattern we stated. One well-known impediment to providing that confirmation is that one cannot use the usual logistic regression model estimated via maximum likelihood to analyze data if a pattern of values invariably produces an outcome, because coefficients march toward infinity in that case (Hosmer and Lemeshow 1989). However, there are other models one can use for such data.

We use Firth's (1993) penalized maximum likelihood (PML) approach to logistic regression estimation as implemented by Heinze and Schemper (2002) in Stata.

Model 1 in Table 6 identified three of the four variables as causal. Technically, this specification is incorrect in that we did not construct Y on the basis of single variables. Thus, we should expect the model to fail. Still, the model does pick out three of the four elements of the causal story. However, the analogous QCA analysis, despite the method's purported ability to identify complex causal recipes on the basis of specification of the separate variables, missed one of two causal configurations while identifying some noncausal factors as causal.

Model 2 specifies the correct causal model and finds that both interactions are discernibly different from zero; model 3, which adds a noncausal variable, correctly identifies the causal factors. Table 6 also contains a model with $X_1 \times X_2$, $X_3 \times X_4$, and $X_1 \times X_2 \times X_4$, which tests (1) whether QCA finds a relationship we overlooked and (2) whether this three-way interaction dominates $X_1 \times X_2$, as QCA implies. We find, however, that coefficients for $X_1 \times X_2$ and $X_3 \times X_4$ are statistically

Table 5. Study 1, QCA Analysis with Equifinal Causes and Noncausal Factor ($n = 40$)

Solution	Outcome $Y = 1$			Outcome $Y = 0$		
	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
Complex	22	$\sim x1 * x3 * x4 * \sim z1$ $x1 * x2 * x3 * \sim x4$ $x1 * x2 * \sim x3 * x4$ $x1 * x3 * x4 * z1$ $x1 * x2 * \sim x4 * \sim z1$ $x1 * x2 * \sim x3 * \sim z1$ $\sim x1 * x2 * x3 * x4$ $x2 * x3 * x4 * z1$.9375, 1.0000	22	$\sim x1 * \sim x4 * \sim z1$ $x1 * \sim x2 * \sim x4$ $\sim x1 * x3 * \sim x4$ $\sim x1 * \sim x2 * \sim x3 * \sim z1$	1.0000, 1.0000
Intermediate	20	$\sim z1 * x4 * x3 * \sim x1$ $x4 * \sim x3 * x2 * x1$ $\sim x4 * x3 * x2 * x1$ $z1 * x4 * x3 * x1$ $\sim z1 * \sim x3 * x2 * x1$ $\sim z1 * \sim x4 * x2 * x1$ $x4 * x3 * x2 * \sim x1$ $z1 * x4 * x3 * x2$.9375, 1.0000	16	$\sim z1 * \sim x4 * \sim x1$ $\sim x4 * x3 * \sim x1$ $\sim x4 * \sim x2 * x1$ $\sim z1 * \sim x3 * \sim x2 * \sim x1$	1.0000, 1.0000
Parsimonious	22	$x4 * z1$ $x1 * x2 * \sim x4$ $\sim x1 * x3 * x4$ $x1 * x2 * \sim x3$ $x1 * \sim x3 * x4$ $x2 * \sim x3 * x4$.9375, 1.0000	22	$\sim x2 * \sim x4$ $\sim x1 * \sim x4$ $\sim x2 * \sim x3$ $\sim x1 * \sim x3$	1.000, 1.000

Table 6. Penalized Maximum Likelihood Logistic Regression Analysis of Study 1 Data of $Y = 1$ and $Y = 0$

Outcome \rightarrow	Model 1		Model 2		Model 3		Model 4		Model 5	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
Constant	-8.735* (3.427)	8.735* (3.427)	-3.891* (1.429)	3.891* (1.429)	-3.545* (1.788)	3.545* (1.788)	-3.888* (1.429)	3.888* (1.429)	-1.490* (.472)	1.490* (.472)
X_1	5.367* (2.332)	-5.367* (2.332)								
X_2	4.392* (2.135)	-4.392* (2.135)								
X_3	1.841 (1.733)	-1.841 (1.733)								
X_4	7.446* (3.251)	-7.446* (3.251)								
$X_1 \times X_2$			6.598* (2.043)	-6.598* (2.043)	6.028* (1.843)	-6.028* (1.843)	6.286* (2.055)	-6.286* (2.055)		
$X_3 \times X_4$			6.455* (2.048)	-6.455* (2.048)	5.889* (1.860)	-5.889* (1.860)	6.449* (2.050)	-6.449* (2.050)	4.033* (1.545)	-4.033* (1.545)
Z_1					-0.043 (1.933)	.043 (1.933)				
$X_1 \times X_2 \times X_4$							-0.793 (2.141)	.793 (2.141)	3.078 (1.627)	-3.078 (1.627)

Note: Values in parentheses are standard errors.
*Coefficient is discernibly different from zero at or below $\alpha = .05$.

significant, while the $X_1 \times X_2 \times X_4$ coefficient is not. The result matches the construction of Y .

Model 5 specifies the result QCA identified, including only the three-way interaction and $X_3 \times X_4$. In this specification, $X_3 \times X_4$ is still statistically significant, while the $X_1 \times X_2 \times X_4$ coefficient is indiscernible from zero. Thus, Table 6 indicates that QCA does not find the correct causal process and that PML logistic regression offers a statistical means for studying such data.⁹ These results, coupled with our knowledge of the construction of Y , make it clear that QCA is wrong.

Finally, note that the $Y = 1$ and $Y = 0$ PML results are equivalent, as only the signs differ. Study 2 explores how QCA finds asymmetric causes, but first we discuss key implications of study 1.

6.2. Discussion of Study 1 Results

6.2.1. Deterministic epistemology. Study 1 uses nonparametric data; specifies deterministic, sufficient, causes; eschews sampling; and imposes equifinality. QCA proponents have identified these conditions as important ontological commitments (e.g., Mahoney 2010; Ragin 2008a; Ragin and Rihoux 2004); thus, study 1 tests the ability of QCA under analytic “best conditions,” as described by QCA proponents. If real-world analysts meet the conditions above and possess data on the full set of causal factors and have all the relevant cases, then study 1 mimics a real-world analysis that completely matches QCA proponents’ conditions, plus the rare state of having full knowledge. Even so, QCA rarely selects the correct causal recipes, raising serious questions about its effectiveness even when real-world conditions match stated QCA assumptions.

At this juncture, the most important issue to address is the foundational assumption of determinism. To proceed, we first distinguish research method and theoretical perspective. Our question is whether the method and/or applied epistemology is or can be deterministic. We argue that sociologists, like physicists, are free to prefer deterministic theories or stochastic theories. For all we know, the social universe may operate with the sociological analog of relativity, such that all proceeds inexorably from origin to a fated end along a path completely determined by first principles, even if humans are still in the process of discovering those principles. This is the Einsteinian position on matter and energy, hence his doubt that God plays dice with the cosmos (Einstein

Table 7. Expected Bias Using Stochastic and Deterministic Epistemologies in Stochastic and Deterministic Worlds

Strategic Epistemology	True Ontological Condition	
	$ \varepsilon > 0$, Stochastic	$\varepsilon = 0$, Deterministic
$ \varepsilon^* $ estimated as ≥ 0 , stochastic	$E(\hat{\varepsilon} - \varepsilon^*) = 0 \rightarrow$ results unbiased	$E(\hat{\varepsilon} - \varepsilon^*) = 0 \rightarrow$ results unbiased
ε^* fixed at 0, deterministic	$E(\hat{\varepsilon} - \varepsilon^*) > 0 \rightarrow$ results biased	$E(\hat{\varepsilon} - \varepsilon^*) = 0 \rightarrow$ results unbiased

Note: ε = true error; ε^* = assumed error; $\hat{\varepsilon}$ = error in estimate.

[1926] 2005). Alternatively, the social universe may operate according to the equivalent of quantum theory, understood as constituted by an irreducibly probabilistic process and thus an impossibility of fully perfect prediction (Gribbin 1984:66). Because the true state of matters is debatable, substantive theories may be either deterministic or probabilistic.

A deterministic ontology is one position; a deterministic epistemology is another. Mahoney (2008) embraces determinism both ontologically and epistemologically. Alas, Table 7 indicates the risky gamble a deterministic epistemology involves. One row reflects a stochastic epistemology, and the other reflects a deterministic epistemology. One column reflects a stochastic ontological reality, and the other reflects a deterministic one. All is well for either epistemology if the world is consistent with its assumption (as shown by the unbiased results on the main diagonal). However, because an epistemological stochasticist estimates an error term, adopting this position in a deterministic world simply produces error estimates of zero. Thus, epistemological stochasticism can produce unbiased results even when its assumption is ontologically incorrect.¹⁰

In contrast, a deterministic epistemology activated in a stochastic world will produce biased estimates because, making no space for error, results will necessarily misallocate any nonzero error, biasing analysis results. Because we are and may forever be ignorant of the true state of the universe, the pragmatic response is an epistemological presumption of stochasticism. Determinist epistemologies entail an unnecessary gamble, a gamble determinists can never confirm they have won nor even discover how much debt the analysis incurs owing to the wager.

6.2.2. *Deterministic methods.* Alas, the gamble can only be rhetorical; regardless of any stated epistemology, no method can instantiate determinism, for a deterministic method requires both omniscient researchers and infinite data. These implications follow because deterministic analyses lack an error term.

The error term in probabilistic models accounts both for the infinite set of unmeasured factors that could partly explain the outcome and for any *irreducibly random* element that may exist. Possible irreducible randomness is not a matter of missing observables in a data set nor such that better, fuller, more accurate measures could conceivably produce an error term of zero; it is, instead, a matter of the possibly mercurial nature of reality such that reality may elude any and all efforts to completely pin it down. Deterministic methods' missing error term means analysts must have sufficient knowledge to (1) rule in or out of the explanation every specific variable, factor, condition, or combination thereof and (2) deny the existence of intransigent randomness.

With respect to the former issue, for example, to study revolutions, analysts would need to know the incidence of agricultural "pests" prior to the revolt; if rare, crops might have higher nutritional value, which might increase peasants' energy and thus facilitate effective protest of usually unprotested grievances. The example concerns just one of the infinity of factors determinists must classify, one by one, as cause or noncause, because assuming $\varepsilon = 0$ necessitates specific classification of every factor that exists. Seawright (2005) suggests this implication.

As for the possibly irreducible randomness that may exist, even if an analyst retreats from a causal aim to an exploratory exercise, as some QCA analysts suggest,¹¹ if the analyst is not certain that the process has absolutely no randomness, and knowing that deterministic explorers cannot sweep any unexamined factors or ineradicably random elements into an error term, the analyst should reach an impasse. He or she should be unable to specify the deterministic model because of incomplete knowledge, whatever the aim.

A deterministic methodologist cannot peek at stochasticists' residuals to confirm that they are zero without implying that stochasticism dominates even in a deterministic universe. Nor can determinists use stochastic models to pare the factors for focus, for a deterministic epistemology is self-contradictory if it requires a deterministic methodologist to use stochastic means to eliminate factors because the deterministic method cannot. Thus, deterministic methods require a researcher to possess an

impossible-to-attain amount of knowledge while providing no route to even begin to attain it. Hence, even if one's theory is deterministic, one requires stochastic methods, or one can have no tractable methods at all. Indeed, even Einstein's (deterministic) theory of relativity, arguably one of humanity's greatest accomplishments, required stochastic methods for its empirical assessment (e.g., Dyson, Eddington, and Davidson 1920).

6.2.3. Deterministic visions, middle positions, and how QCA works. QCA, therefore, necessarily, retreats from determinism, or it would usually reject all causal claims. If the method were truly deterministic, any consistency score short of 1.0 would provide decisive evidence that the configuration is not causal. And, as Lieberman (1994) notes, as sample size increased, so would the inability to identify causal factors, a perverse result of deterministic methods: This far from omniscience, more information means less ability to discern cause. Thus, most deterministic studies would eliminate all theorized causal factors (as we found for pre-STS-51-L blow-by). Hence, consistently deterministic studies would place their authors in the odd position of simultaneously asserting that chance does not exist while repeatedly documenting that none of the theorized causes actually cause the outcome either. With deterministic study after deterministic study reaching this conclusion, determinists would find themselves cumulatively, implicitly documenting the power of the very randomness their epistemological position denies.

Ragin (2000) articulates a middle position that references "almost necessary cause" and "almost sufficient cause." Other analysts also relax the strict demands of necessary and sufficient causation and its accompanying determinism (e.g., Clark, Gilligan, and Golder 2006:313–14). We find such qualifications unhelpful, because there really is no epistemologically coherent middle position between stochasticism ($|\varepsilon| \geq 0$) and determinism ($\varepsilon = 0$).

Efforts to salvage necessary and sufficient causation by qualifying the terms resembles the many efforts to salvage the possibility of an objective science by changing the definition of the word *objective* from meaning apprehension of objects as they themselves are (Kant [1781] 2008)—an impossibility—to something else (e.g., Fuchs 1997). Just as we wonder why some cannot accept the impossibility of objectivity and embrace the accessible gains of systematicity, we wonder why some resist the impossibility of deterministic analysis and eschew the

accessible gains of stochastic epistemologies, especially as embracing epistemological stochasticism does not force one to accept ontological stochasticism.

An inherently dichotomous choice presents itself; either one specifies an error term (or its functional equivalent) or one does not. QCA, by allowing less than 100 percent consistency, implicitly allows an error term (i.e., some factor other than explicitly identified X's causes the outcome). Yet, even with consistency set at 100 percent, determinists must be omniscient to make $\varepsilon = 0$. Thus, QCA cannot be practicable and deterministic. Still, rhetoric around QCA is often deterministic.

QCA's operational retreat from determinism allows analysis to proceed but is insidious. It allows analysts to claim perfect causality without meeting that exacting standard, such that they may mislead themselves and others about the strength of their evidence. And it allows analysts to use a rhetoric of determinism while buffering them from the dire consequences of their claims.

The disjuncture between how QCA is understood to operate and how it actually operates hinders analysts' efforts to use QCA to clarify the epistemological debate. Suffice it to say, we reject deterministic methods—as all analysts must—as well as deterministic epistemological rhetoric—as all analysts should.

6.3. Implications of Study 1 Results for Remaining Analyses

Because we have shown that QCA is not deterministic, we no longer construct data using deterministic methods. Relaxing this constraint simplifies data construction and increases the flexibility of tests we can apply.

7. STUDY 2: INTERROGATING ASYMMETRIC CAUSALITY

7.1. Study 2 Methods

In Study 1, 10 of 12 QCA solutions were incorrect. Curiously, each correct solution for $Y = 0$ was paired with an incorrect solution for $Y = 1$, a possibility given QCA strategy for studying asymmetric causality. Study 2 interrogates the QCA strategy for studying asymmetric causality.

The advised QCA procedure for assessing asymmetric causality is to analyze $Y = 1$ and $Y = 0$ (or $Y = \text{high}$ and $Y = \text{low}$) separately (De Meur and Rihoux 2002; Ragin 2008a:137–38). We followed QCA advice.

Our aim was to see whether QCA will find asymmetry even with symmetric causes; to ensure symmetry, we built the data using a parametric model.¹²

7.2. Study 2 Results

We present the complex and parsimonious solutions, highlighting the complex solution (see Table 8). Symmetry might be difficult to discern, but one sign would entail a recipe of, for example, $X_1 = 1$, $X_2 = 0$ for $Y = 1$ and $X_1 = 0$, $X_2 = 1$ for $Y = 0$. However, we find no clear evidence of symmetry, finding four causal recipes of $Y = 1$ and three causal recipes of $Y = 0$ across the solutions.

One might contend that this particular data set is rife with asymmetries, which would explain our inability to locate any evidence of symmetric causality. The problem, however, is that we know the data were constructed using a symmetric process.¹³ Thus, any claim of asymmetry is incorrect, raising major questions about the method's ability to discern asymmetric causality.

Closer inspection reveals that this strategy of analyzing both $Y = 1$ and $Y = 0$ is the root of the problem. If one follows this strategy, the two analyses will divide all causal configurations into three categories, with λ and τ as consistency thresholds:

Category A: $Y = 1$ occurs in more than λ percent of the cases in the causal configuration.

Category B: $Y = 0$ occurs in more than τ percent of the cases in the causal configuration.

Category C: $Y = 1$ occurs in less than or equal to λ percent and $Y = 0$ occurs in less than or equal to τ percent of the cases in the causal configurations.

Analysts are advised that the minimum consistency threshold, expressed as a percentage, is 75 (Ragin 2008a). Using percentages, if one sets $\lambda + \tau > 100$ (as counseled), then category C may contain one or more configurations. Because analyses of $Y = 1$ and $Y = 0$ both set aside category C, the existence of this category masks any symmetry embedded in the causal process.

To see the implications of this fact, we reanalyzed the study 2 data using the truth table in Table 9. The last column lists the number of

Table 8. Study 2, Results for Analyses of $Y = 1$ and $Y = 0$ for CG2 Data, Consistency Threshold of .8 ($n = 20$)

Solution	$Y = 1$			$Y = 0$		
	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
	One-case Threshold			One-case Threshold		
Complex	13	$x2^* \sim x4$ $\sim x1^* x2$ $x1^* \sim x3^* \sim x4$.5385, 1.0000	13	$x1^* \sim x2^* x3$ $\sim x1^* \sim x2^* \sim x3^* x4$.4286, 1.0000
Parsimonious	13	$x1^* \sim x3$ $x2^* \sim x4$ $\sim x1^* x2$.5385, 1.0000	13	$x1^* \sim x2^* x3$ $\sim x2^* \sim x3^* x4$.4286, 1.0000

Table 9. Study 2 Truth Table for Analyses of $Y = 1$ and $Y = 0$, CG2 ($n = 20$)

Causal Configuration	X_1	X_2	X_3	X_4	$Y = 1$	$Y = 0$	Count
A	0	0	0	0	.67	.33	3
B	1	0	0	0	1.00	.00	1
C	0	1	0	0	1.00	.00	1
D	1	1	0	0	1.00	.00	1
E	0	0	1	0	—	—	0
F	1	0	1	0	.00	1.00	1
G	0	1	1	0	1.00	.00	1
H	1	1	1	0	1.00	.00	1
I	0	0	0	1	.00	1.00	1
J	1	0	0	1	—	—	0
K	0	1	0	1	1.00	.00	1
L	1	1	0	1	—	—	0
M	0	0	1	1	.33	.67	3
N	1	0	1	1	.00	1.00	1
O	0	1	1	1	1.00	.00	1
P	1	1	1	1	.75	.25	4

cases that fall into each causal configuration. Rows E, J, and L have no observations and thus are eliminated for failing to meet the frequency threshold of 1. The $Y = 1$ column indicates the proportion of cases in the causal configuration for which $Y = 1$; the $Y = 0$ column reports the analogous proportion for $Y = 0$.

The recommended analytic approach appears in the first two panels of Table 10. Columns 5 and 6 indicate which causal configurations are specified to produce outcome $Y = 1$ or $Y = 0$ and the causal configurations implicitly claimed to not produce the outcome. Thus, an analyst investigating $Y = 1$ specifies seven causal configurations—B, C, D, G, H, K, and O—as causal. In doing so, the analyst implicitly claims that the remaining configurations do not produce $Y = 1$. In contrast, an analyst studying $Y = 0$, using the recommended procedure, specifies only three configurations—F, I, and N—as causal. By implication, therefore, this analyst specifies that the remaining configurations do not produce the outcome $Y = 0$. Note that three causal configurations—A, M, and P—are never included as possible determinants of Y . We contend that it is the neglect of AMP that is feeding back to produce distortion that generates a report of asymmetric causality.

If we presume that asymmetric causation is possible, and thus causes of $Y = 1$ do not constrain causes of $Y = 0$, then if we specify that AMP

Table 10. Study 2, Results for Analyses of Asymmetric Causality for CG2 Using Alternative Strategies of Configuration Assignment, One-case Threshold ($n = 20$)

Outcome/Solution	R	Causal Recipes	Coverage, Consistency	Specified Expectation of Causal Configurations	Consistency Threshold
<i>Recommended analytic approach</i>					
<u>Panel 1, $Y = 1$</u> Complex	13	$x2^* \sim x4$ $\sim x1^* x2$.5385, 1.0000	$Y = 1$ $\overline{B, C, D, G, H, K, O}$	$Y \neq 1$ $\overline{A, F, I, M, N, P}$.75
Parsimonious	13	$x1^* \sim x3^* \sim x4$ $x1^* \sim x3$ $x2^* \sim x4$ $\sim x1^* x2$.5385, 1.0000	B, C, D, G, H, K, O	A, F, I, M, N, P .75
<u>Panel 2, $Y = 0$</u> Complex	13	$x1^* \sim x2^* x3$ $\sim x1^* \sim x2^* \sim x3^* x4$.4286, 1.0000	$Y = 0$ $\overline{F, I, N}$	$Y \neq 0$ $\overline{A, B, C, D, G, H, K, M, O, P}$.75
Parsimonious	13	$x1^* \sim x2^* x3$ $\sim x2^* \sim x3^* x4$.4286, 1.0000	F, I, N	$A, B, C, D, G, H, K, M, O, P$.75
<i>Mutually exclusive specification of configurations</i>					
<u>Panel 3, $Y = 0$</u> Complex	13	$\sim x1^* \sim x2^* \sim x3$ $x1^* \sim x2^* x3$ $x1^* x3^* x4$ $\sim x1^* \sim x2^* x4$ $\sim x2^* x3^* x4$	1.000, .5385	$Y \neq 0 \rightarrow Y = 1$ $\overline{B, C, D, G, H, K, O}$	$Y = 0$ $\overline{A, F, I, M, N, P}$.25
Parsimonious	13	$\sim x1^* \sim x2$ $\sim x2^* \sim x3$ $x1^* x4$	1.000, .5385	B, C, D, G, H, K, O	A, F, I, M, N, P .25

Note: Causal configuration table: $A = 0000, B = 1000, C = 0100, D = 1100, E = 0010, F = 1010, G = 0110, H = 1110, I = 0001, J = 1001, K = 0101, L = 1101, M = 0011, N = 1011, O = 0111, P = 1111$. E, J, and L are always counterfactual configurations, with no observed cases.

does not cause $Y = 1$, what justifies the switch to presuming that AMP is noncausal when we move to the analysis of $Y = 0$? Although this assumption is defensible, given the claim that AMP is not causal for $Y = 1$ and a belief in a possibility of asymmetric causation as defined in QCA, should one not at least specify a provisional $Y = 0$ analysis that explores whether $Y = 0$ is produced by AMP? Panel 3 specifies just such an exploratory analysis. Implementing this specification as the companion of the $Y = 1$ analysis violates recommended QCA procedure but fulfills a deeper principle of analytic completeness by exploring the possible causal implications of all observed causal configurations.¹⁴ Thus, no configuration is ruled out of contributing to some outcome.¹⁵

Summary statistics for panel 3 are the opposite of those obtained in panel 1, suggesting that explanations for $Y = 1$ and $Y = 0$ are essentially the same. Yet causal recipes for $Y = 1$ and $Y = 0$ are different and not clearly opposite. There are two signs of symmetry, but they are opaque. First, the complex solution contains $x_1 \sim x_3 \sim x_4$ for $Y = 1$ and $x_1 \cdot x_3 \cdot x_4$ for $Y = 0$. Thus, once we set aside x_1 , the role of $x_3 \cdot x_4$ is opposite for the two outcomes. Second, for $Y = 1$, the parsimonious solution contains $\sim x_1 \cdot x_2$, and for $Y = 0$, it contains $\sim x_1 \sim x_2$. Thus, with the proviso that $\sim x_1$ is required, the role of x_2 in producing $Y = 1$ is just opposite its role in producing $Y = 0$. But these two sets of recipes offer a minimal sign of the symmetry embedded deep in the results.

To excavate that symmetry Table 11 traces the implications of the causal recipes from the complex solutions of panels 1 and 3 of Table 10. The first column contains causal recipes. In column 2, which translates the causal recipe into X_1 , X_2 , X_3 , and X_4 , zeros indicate nonmembership in the positionally identified set, ones signify membership, and dashes indicate the variable is not relevant for the causal recipe. When the factor is irrelevant, replacing the dash with either a zero or a one is equally correct. Column 3 provides just such a set of substitutions, expanding directly the implications of the recipes to reveal the actual causal configurations they summarize.

Column 3 indicates that some implications of a causal recipe duplicate the implications of some other causal recipe. Thus, the three causal recipes for $Y = 1$ imply 10 causal configurations, but there really are only 7 nonredundant causal configurations.

The key conclusion is that when we compare the causal recipes for $Y = 1$ with those for $Y = 0$, it appears that asymmetry holds. However, Table 11 reveals that a deeper symmetry is hidden by the recipes: The

Table 11. Study 2, Implications of Panel 2 Causal Recipes for $Y = 1$ and $Y = 0$, Complex Solution, CG2 Data

	Translation into X_1, X_2, X_3, X_4	Implications of Translation	Summary of the Disaggregation
Causal recipes for $Y = 1$			
$x_2^* \sim x_4$	— 1 — 0	0100 == C ($Y = 1$ dup 1) 1100 == D ($Y = 1$ dup 2) 0110 == G ($Y = 1$ dup 3) 1110 == H	7 nonredundant causal configurations of $Y = 1$ implied in the 3 causal recipes
$\sim x_1^* x_2$	0 1 — —	0100 == C ($Y = 1$ dup 1) 0101 == K 0110 == G ($Y = 1$ dup 3) 0111 == O	
$x_1^* \sim x_3^* \sim x_4$	1 — 0 0	1000 == B 1100 == D ($Y = 1$ dup 2)	
Causal recipes for $Y = 0$			
$\sim x_1^* \sim x_2^* \sim x_3$	0 0 0 —	0000 == a 0001 == i ($Y = 0$ dup 1) 1010 == f	6 nonredundant causal configurations of $Y = 0$ implied in the 5 causal recipes Implications in boldface type were configurations implicated in the recommended analysis of $Y = 0$ (0001 was implied by the 4-way interaction identified in Table 10, panel 2)
$x_1^* \sim x_2^* x_3$	1 0 1 —	1011 == n ($Y = 0$ dup 2)	
$x_1^* x_3^* x_4$	1 — 1 1	1011 == n ($Y = 0$ dup 2) 1111 == p	
$\sim x_1^* \sim x_2^* x_4$	0 0 — 1	0001 == i ($Y = 0$ dup 1) 0011 == m ($Y = 0$ dup 3)	
$\sim x_2^* x_3^* x_4$	— 0 1 1	0011 == m ($Y = 0$ dup 3) 1011 == n ($Y = 0$ dup 2)	
All 13 observed causal configurations are causal There is no asymmetric content: the solution to $Y = 1$ implies the solution to $Y = 0$			
Conclusion 1			
Conclusion 2			

Note: Causal configuration table: $A = 0000$, $B = 1000$, $C = 0100$, $D = 1100$, $E = 0010$, $F = 1010$, $G = 0110$, $H = 1110$, $I = 0001$, $J = 1001$, $K = 0101$, $L = 1101$, $M = 0011$, $N = 1011$, $O = 0111$, $P = 1111$. E, J, and L are always counterfactual configurations, with no observed cases.

causal pathways are such that knowing the causal recipes (or, at least, their implied causal configurations) for one outcome implies the causal recipes for the other. The final column of Table 11 reveals this to be the case: once we know the recipes for $Y = 1$, we need not analyze $Y = 0$, for its causal recipes are a perfectly symmetric set (save for the counterfactual configurations about which one, truly, can say nothing).

This conclusion becomes apparent, however, only through the analysis presented in panel 3 of Table 10. If one conducts the advised analyses and sets $\lambda + \tau > 100$, as in panel 2 of Table 10, then symmetry will likely vanish. This is noted by the recipes, translation, and implications in boldface type in Table 11; these are the results obtained in panel 2 of Table 10. The reason those results are only a subset of those obtained in panel 3 is that in the panel 2 analysis, several causal configurations claimed not to cause $Y = 1$ are eliminated from consideration of possibly causing $Y = 0$ *a priori*. This is the approach Ragin (2008a) recommends, and if one follows that recommendation one will be able to obtain results that reflect burgeoning asymmetry. Yet it appears the asymmetry obtained is a spurious result of QCA.

7.3. Discussion of Study 2 Results

7.3.1. Omitting cells to produce incomplete comparisons. The neglect of causal configurations (e.g., AMP above) extends QCA advice for the two-by-two table to the m-way table. We offer U.S. presidential elections to illustrate costs of the ignore-a-cell approach. Between 1871 and 2009 there were 165 presidential party nominees and certified independent general election candidates. Thirty-five presidential candidates won; all but 1 were white. Ninety-nine white and 31 black candidates lost.

We wonder, do victors share the same race? The ignore-a-cell approach compares cells (0,1) and (1,1), finding 1 black and 34 white victors. Whites are far more likely to win than are blacks. We conclude that victors share the same race; being white seems to be a cause of electoral victory.

We wonder, does being white matter? The ignore-a-cell approach compares cell (1,1)'s 34 entries (white victors) to cell (1,0)'s 99 entries (white also-rans). It appears that whites are much more likely to lose than to win. We conclude being white seems to be a cause of electoral defeat.

These conclusions might be viewed as revealing asymmetric causation (Ragin 2008a). We contend the conclusions are contradictory and

irreconcilable without information from the missing cell, for with those data, we can fully compare white and black electoral prospects. Upon consideration of cell (0,0), we find that although only 26 percent of white candidates win the presidency, fewer than 4 percent of black candidates win. Thus, considering the full table reveals that one finding is wrong: being white is positive for candidates' presidential prospects, not negative.

Incorrect and contradictory findings often follow the neglect of a table cell because, as Clark et al. (2006:320) note, causal claims are conditional claims. It is odd to claim otherwise. The oddity is that $\Pr(Y = 1|X = x)$ fully determines $\Pr(Y = 0|X = x)$, so there is no more information to extract by analyzing $Y = 0$: If we know $\Pr(Y = 1|X = x) = .8$, then we know $\Pr(Y = 0|X = x) = .2$. In contrast, absent information on the marginal distributions, cell (0,0) contains nonredundant information. Yet QCA analysts are discouraged from accessing it.

7.3.2. Defining asymmetric causation. In QCA, different causes of $Y = 0$ and $Y = 1$ (or $Y = \text{high}$ and $Y = \text{low}$) reflect asymmetric causation, and QCA may find asymmetry because QCA can find necessary or sufficient causation (Ragin 2008a:17–20) using the ignore-a-cell approach (Ragin 2000).

We distinguish two kinds of asymmetry. In “functional form asymmetry,” the difference in Y when $X = 1$ is compared with $X = 0$ differs from the difference in Y when $X = -1$ is compared with $X = 0$; that is, the absolute value of the results of “moving” from some center point on X differ depending on the direction moved (0 moved to 1 vs. 0 moved to -1). Yet the absolute value of the results of “moving” between two points are equal regardless of the direction moved (0 moved to 1 vs. 1 moved to 0). Functional form asymmetry is a basic hypothesis statistical analysts may explore using higher order terms, asymmetric link functions (e.g., complementary log-log), or in other ways, such as piecewise regression and locally weighted scatterplot smoothing regression (Cleveland 1979).

We view necessity and sufficiency as referencing functional form asymmetry, reasoning as follows. First, assume dichotomous Y and trichotomous X , with $X = 3$ a necessary cause of $Y = 1$ (i.e., if $X \leq 2$, $Y = 0$, but if $X = 3$, $Y = 1$). Thus, all of the change in Y occurs in the move from $X = 2$ to $X = 3$. Valid studies will show that

$$|\Pr(Y = 1|X = 3) - \Pr(Y = 1|X = 2)| \neq |\Pr(Y = 1|X = 2) - \Pr(Y = 1|X = 1)|, \quad (3.1)$$

but

$$|\Pr(Y = 1|X = 3) - \Pr(Y = 1|X = 2)| = |\Pr(Y = 1|X = 2) - \Pr(Y = 1|X = 3)|. \quad (3.2)$$

Sufficient causation could be similarly expressed. Equations (3.1) and (3.2) reflect functional form asymmetry; hence, necessary causation and sufficient causation reflect functional form asymmetry.

Another type of asymmetry refers to time, à la the English dominance example Lieberman (1985) offered in a chapter titled “Asymmetrical Forms of Causation.” The example considers the rising dominance of the English language over the course of the twentieth century, dominance that might be partly explained by the pre-1939 power of Britain and the postwar rise of the United States. Yet even if those causes were undone (e.g., the United States drowns in rising debt), English may remain dominant. Reversing the causes need not reverse their effects.

We term this kind of asymmetry “dynamic asymmetry” because it is inaccessible without cross-time data. Dynamic asymmetry is not a functional form issue. To clarify with an example, assume that an autumn survey of college sophomores ascertained the number of hours they studied a given week that term and the number of classes in which they enrolled. We find students with two classes studied 5 hours a week and students with four classes studied 20 hours a week. In a spring resurvey, we find that students who formerly had two classes and now have four classes now study 20 hours a week. We also find students who had four classes the previous term and now have two classes also study 20 hours a week. However, students with two (four) classes both terms studied 5 (20) hours a week in both terms. In dynamic asymmetry, the direction of change between two points matters. In the example, moving a student from taking two classes to four classes raises her studying by 15 hours a week, but moving a student from taking four classes to two leaves her studying time unchanged.

In equations, functional form asymmetry is such that

$$|E[Y_{X=0}] - E[Y_{X=1}]| \neq |E[Y_{X=0}] - E[Y_{X=-1}]| \quad (4.1)$$

but

$$|\Delta E[Y_{X=0} \rightarrow Y_{X=1}]| = |\Delta E[Y_{X=1} \rightarrow Y_{X=0}]|, \quad (4.2)$$

while dynamic asymmetry is such that

$$|\Delta E[Y_{X=0} \rightarrow Y_{X=1}]| \neq |\Delta E[Y_{X=1} \rightarrow Y_{X=0}]|,^{16} \quad (5)$$

where the arrow (\rightarrow) signifies the direction of change in X .

Our example of dynamic asymmetry is a form of lock-in: once exposed to a need to study 20 hours, students allocate 20 hours to study even when they take fewer courses. Lock-in is a familiar form of dynamic asymmetry: English dominance, the QWERTY keyboard, driving on the right side of the road in the United States, substance addiction, and other social phenomena collected under the concept of habit variously defined (Camic 1986) are all examples. Notably, QCA's technique for finding asymmetric causation does not access dynamic asymmetry and does not effectively access functional form asymmetry either.

7.3.3. The pivotal role of information. Lieberman's (1985) example introduced information on event sequencing. Narrative analysis (Mahoney 1999) or process-tracing (George and Bennett 2005) also rely on additional often nonstandard information. Some approaches to investigating sequencing (Abbott and Hrycak 1990) or duration (Wu 2003) might also be useful. There may be other possibilities. The QCA strategy, in contrast, creates a serious identification problem (Manski 1995) and attempts to solve it by eliminating rather than supplementing the information available.

Consider a social phenomenon and its reverse, such that

$$\Pr(Y = 1) = f(\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \varepsilon_j), \quad (6)$$

and

$$\Pr(Y = 0) = f(\gamma_0 + \gamma_1 X_{1j} + \gamma_2 X_{2j} + \gamma_3 X_{3j} + \delta_j), \quad (7)$$

with f a not necessarily parametric function and errors (ε , δ) equal to zero if determinism holds. Moving from expression (6) to (7) reverses the dependent variable, just as QCA advocates. However, as Grofman and Schneider (2009:669) assert, $\gamma_0 = -\beta_0$, $\gamma_1 = -\beta_1$, $\gamma_2 = -\beta_2$, and $\gamma_3 = -\beta_3$. The proof of this transmethodological conclusion can begin by assuming equations (8.1) and (8.2):

$$\Pr(Y_i = 1) = e^{bx} / (1 + e^{bx}), \quad (8.1)$$

and

$$\Pr(Y_i = 0) = e^{gx} / (1 + e^{gx}). \quad (8.2)$$

Thus,

$$1 = \Pr(Y_i = 1) + \Pr(Y_i = 0), \quad (8.3)$$

which implies

$$1 = (e^{bx} / (1 + e^{bx})) + (e^{gx} / (1 + e^{gx})). \quad (8.4)$$

Simplifying,

$$1 = (e^{bx} + 2e^{bx+gx} + e^{gx}) / [(1 + e^{bx})(1 + e^{gx})]. \quad (8.5)$$

Solving for b , most terms cancel, reducing to

$$b = -g. \quad (8.6)$$

An asymmetric link function for $\Pr(Y = 1)$ requires its opposite link function for $\Pr(Y = 0)$ to test whether $\beta = -\gamma$. The test would show the problem remains.

In short, to separately analyze the reverse of a dependent variable's coding is to attempt to distinguish different absolute values of a variable's effect on two perfectly collinear outcomes. Absent introduction of new information, this cannot be done.

New information is key. Assumptions are one kind of new information one can invoke to identify unidentified parameters. However, estimates identified via strong assumptions contain no content independent of the assumptions invoked. In specifying $\lambda + \tau > 100$, analysts impose different, strong, and contradictory assumptions on analyses of $Y = 1$ and $Y = 0$, which can easily mask the fundamental symmetry of the causal structure of $Y = 0$ and $Y = 1$. Thus, the pairing in study 1 of a correct $Y = 0$ solution with an incorrect $Y = 1$ solution is revealed to flow from the invocation of contradictory assumptions that superficially resolve identification problems. The implication is that there is no asymmetric content in QCA results beyond that imposed by assumption.

7.4. *Implications of Study 2 Results for Remaining Analyses*

One important finding is that there is no value to analyzing both $Y = 1$ and $Y = 0$ (or $Y = \text{high}$ and $Y = \text{low}$). As we continue, therefore, we will analyze only $Y = 1$ (or $Y = \text{high}$).

8. STUDY 3: IDEAL DATA, YET QUALITATIVE COMPARATIVE ANALYSIS FAILS?

8.1. *Does QCA Have a Consequential Unstated Assumption?*

Study 1 data were designed to match the stated epistemological assumptions of QCA proponents: nonparametric, deterministic data reflecting sufficient, equifinal causes measured on the full population. Yet QCA failed. Study 3 explores whether QCA's failure is explained by an unstated QCA assumption both Lieberman (1991) and Tilly (1984) posited for Mill's methods.

In study 1 (and in reality), an outcome can be overdetermined; all X 's for a case can equal 1 even if subsets of $X_k = 1$ are sufficient for $Y = 1$. The social world is rife with such situations. There may be two sufficient reasons for "Joan's" hiring: (1) her 4.0 grade point average and her physics degree on one hand and (2) her mother owning the company on the other. As another example, there may be three constellations of reasons why prospective students are admitted to elite colleges: (1) academic excellence (measured by grades and scholarly honors), (2) athletic excellence (measured by varsity letters) combined with academic adequacy (measured by the lack of grades below C+), and (3) parental wealth (measured by income, liquid assets, and illiquid assets) coupled with a high school diploma or equivalent. Some applicants (e.g., valedictorian varsity quarterbacks) will satisfy multiple causal constellations (e.g., both 1 and 2 above), making their admission overdetermined. A "no overdetermination" assumption denies such possibilities. Thus, we ask, Does QCA deny the possibility of overdetermination despite its common real-world occurrence?

To assess how QCA handles overdetermined data, we deleted the three overdetermined cases from the data used in study 1. Because QCA produced incorrect causal solutions when the overdetermined cases were present, if we obtain the correct causal solution after removing the overdetermined cases, it would appear that QCA implicitly assumes that

Table 12. Study 3, Overdetermined Cases Eliminated, Analysis of $Y = 1$, Crisp Group Data, Consistency Threshold of .8, One-case Threshold ($n = 37$)

Solution	R	Causal Recipes	Coverage, Consistency
Complex	13	$x1*x2*\sim x4$ $x1*x2*\sim x3$ $\sim x2*x3*x4$ $\sim x1*x3*x4$	1.0000, 1.0000
Intermediate	13	$x1*x2*\sim x4$ $x1*x2*\sim x3$ $\sim x2*x3*x4$ $\sim x1*x3*x4$	1.0000, 1.0000
Parsimonious	13	$x1*x2$ $x3*x4$	1.0000, 1.0000

overdetermination is not possible. Table 12 contains the results of our analysis.

8.2. *QCA Results after Deletion of Overdetermined Cases*

Curiously, QCA still produces an incorrect complex solution, repeatedly finding that irrelevant factors matter. For example, QCA finds that X_4 is relevant for whether the X_1, X_2 combination is causal. By construction, we know that this is incorrect. We find, therefore, that all four causal recipes in the complex solution are wrong.

However, QCA finds the correct solution with its parsimonious solution, when formerly the parsimonious solution for $Y = 1$ was incorrect. This result, however, far from providing a basis for accepting QCA, raises serious doubts about the method.

8.3. *Discussion of Study 3 Results*

8.3.1. *Implications of QCA's (partial) success when overdetermined cases are deleted.* Deleting overdetermined cases does not make QCA find a correct complex solution. Yet because deleting overdetermined cases makes QCA produce the correct parsimonious solution, we find that a hidden “no overdetermination” assumption can partly explain QCA’s failure on data that match proponents’ stated ontological commitments. This result has three key implications.

First, if multiple sets of factors can cause the outcome (i.e., if equifinality is possible), then some persons, organizations, nation-states, or

events may have multiple aligned sets of factors, overdetermining the outcome. Indeed, it does not take much imagination to conjure up countless examples of overdetermination to add to our illustrative cases of Joan and elite college admissions. Yet the results of study 3 indicate that to use QCA on a full data set, one must deny overdetermination in the domain under study, despite its common real-world occurrence.

Second, partially repairing QCA by deleting overdetermined cases is a self-contradictory solution to the problem the assumption creates. To delete overdetermined cases, one must already know the composition of each set of necessary and sufficient causes. But if one knows the sets of necessary and sufficient causes already, what is the point of using QCA?

QCA's implicit "no overdetermination" assumption violates a key methodological principle, to wit, that a method should not require one to know the answer to the question one deploys the method to answer. Of course, all methods have assumptions. Yet all assumptions are not equal. For example, in structural equation models, estimating a causal path from A to B could be viewed as assuming a causal path from A to B. But model results allow one to evaluate this path, so this is really a hypothesis, not an assumption. As another example, we assume an error term in ordinary least squares (OLS) regression, but one typically does not estimate the model to learn about the errors, so one has not assumed an answer to the question one uses the method to answer. Many assumptions are either hypotheses or answer questions other than the question the method is deployed to answer.

QCA users ask, "What is causal?" A possible answer to that question is "two different sufficient conditions." Given that a case could satisfy two different sufficient conditions, a solid method must distinguish two different sufficient conditions from one complex condition. QCA appears to solve this estimation problem by denying overdetermination, which, rather than distinguishing the possibilities, extinguishes one possibility.

Third, the meta-analytic implication of study 3 is that simulations are essential for assessing methods. Using simulations, we discovered an implicit QCA assumption that would be extremely difficult to discern if analysts confine their use of the method to real data. Thus, study 3 demonstrates the potential value of simulations for QCA specifically and other methods too.

In sum, Ragin (1987:98) suggests that QCA will work with overdetermined outcomes. But our results indicate otherwise.

9. STUDY 4: SEARCHING FOR CALIBRATION STABILITY AND REPLICABILITY IN FUZZY-SET QUALITATIVE COMPARATIVE ANALYSIS

Another key question is whether QCA results replicate; we address this issue using fsQCA. Liebersson (2004) argues that QCA capitalizes on chance perturbations, which would make replication elusive. Replication may seem of little interest to case-oriented researchers; if the aim is to analyze specific, intrinsically interesting cases, then one might view replication as unnecessary. However, if results change dramatically with small calibration changes or sample changes, one might wonder how robust fsQCA is. Thus, replicability issues are of broad interest.

9.1. *Study 4 Methods*

We constructed a 40-case population using a parametric model reflecting a known set of causal relations and then randomly allocated 20 cases to fuzzy group 1 (FG1) and 20 cases to fuzzy group 2 (FG2). We used a parametric causal structure because it is simpler than the alternative and should be accessible to fsQCA, a method designed for use with continuous variables. Note that because the cases were constructed using coefficients that apply to every case, despite their later placement in different data sets, their causes are the same.

A key part of fsQCA analysis entails calibration. We used multiple calibrations and two calibration logics, empirical calibration and range calibration, explicated in Table 13. We know that few substantive researchers would use multiple calibrations, but we did so in our methodological investigation to decrease the chance that a failure to replicate the known causal structure would be attributed to poor calibration. And, to facilitate a test of stability case-oriented researchers might find germane, we used a strategy that allows us to compare results for one group using different calibration scores and across groups with the same scores.

9.2. *Study 4 Results*

Table 14 contains the calibrations, while Table 15 contains the analysis results. The first issue to consider is whether fsQCA was able to

Table 13. Calibration Strategies

Strategy	Description
Empirical calibration	A given group’s 5th, 50th, and 95th percentiles anchor the thresholds of full nonmembership, crossover, and full membership in the set. One may obtain an empirical calibration for any continuous variable. For example, for earnings, the empirical calibration may provide a midpoint around \$40,000.
Range calibration	The 5th percentile, midpoint, and 95th percentile of the logical range of the variable anchors the thresholds of full nonmembership, crossover, and full membership in the set. For example, for earnings, the range calibration may provide a midpoint around \$50,000,000.

Note: Although our variables do not have ranges that are so large, anchor values may differ across the calibration logics.

identify the correct causal recipes. The raw data, before fsQCA calibration, match equation (9):

$$Y_i = .613 - .004 \times X_{1i} + .030 \times X_{2i} - .029 \times X_{3i} + .010 \times X_{4i} + \varepsilon_i. \tag{9}$$

The R^2 value from the OLS regression is .981, signifying that the error variance, σ^2_ε , is very small. A purportedly deterministic method should be more likely to succeed for processes with high R^2 values. Coefficients for the constant, X_2 , and X_3 have p values below .0005; other p values exceed .57. No interaction effects were used in constructing Y . Even so, every causal recipe in the complex solutions identifies two- or three-way interactions. The parsimonious solutions fare better, with the majority of the recipes referencing single causes. Yet even when single causes are referenced, they are more often wrong than right. Thus, even though an OLS regression model can explain 98 percent of the variance, fsQCA consistently fails to identify the correct causal factors.

Looking more closely, we consider the range calibration analysis first. None of the causal recipes match across FG1 and FG2. In fact, one recipe is exactly the opposite in the two data sets: A causal recipe for $Y = 1$ in FG1 is $X_2 = X_3 = 1$, while a causal recipe for $Y = 1$ in FG2 is $X_2 = X_3 = 0$. These results indicate a failure to replicate.

Table 14. Four Sets of Calibration Anchors for Study 4

Calibrations	Variables and Thresholds					
	X_1 out	X_1 middle	X_1 in	X_2 out	X_2 middle	X_2 in
1. G1 data, range	.0500000	.5000000	.9500000	1.00000	10.00000	19.00000
2. G1 data, empirical	.0262917	.4986896	.9362063	1.86918	13.15776	19.51347
3. G2 data, empirical	.0600459	.4423967	.9710752	1.72417	9.633014	17.55905
4. All data, empirical	.0125181	.4986896	.9775083	1.009517	11.75234	19.51347
	X_3 out	X_3 middle	X_3 in	X_4 out	X_4 middle	X_4 in
1. G1 data, range	1.00000	10.00000	19.00000	.0500000	.5000000	.9500000
2. G1 data, empirical	2.970067	11.19308	17.32408	.2269875	.5228702	.9442348
3. G2 data, empirical	2.347759	9.72293	19.62373	.0320468	.553797	.9623687
4. All data, empirical	1.645022	10.34501	19.29447	.0373015	.5307223	.978161
	Y out	Y middle	Y in			
1. G1 data, range	.0500000	.5000000	.9500000			
2. G1 data, empirical	.2753639	.739311	.9708401			
3. G2 data, empirical	.2757255	.4770155	.7160342			
4. All data, empirical	.2553267	.6818468	.9708401			

Table 15. Study 4, Results for Two Fuzzy Groups of 20 from Fuzzy Simulated Data Set of $n = 40$, $Y = 1$ Analyses

Solution	Coverage, Consistency		R	Coverage, Consistency		
	Causal Recipes	Analyses of FG1 Data		Causal Recipes	Analyses of FG2 Data	
	One-case Threshold, Consistency Threshold of .8			One-case Threshold, Consistency Threshold of .8		
Range calibration using FG1 Complex	11	~x1*x2 ~x3*x4 x2*x3 x1*x4	.8763, .9028	12	~x2*~x3 x1*~x3 x1*~x2 x2*x4 ~x1*x2*x3 All configurations meet frequency threshold	.8896, .8256
	16	~x1*x2 ~x3*x4 x2*x3 x1*x4 x2 x4	.8763, .9028			
Parsimonious	11		.9413, .8624		All configurations meet frequency threshold	
Empirical calibration using FG1 data Complex	11	~x1*~x3 ~x3*x4 ~x1*x2*x4	.8016, .8474	12	x1*x2*~x3 x2*~x3*x4 x1*x2*x4	.6152, .9268

(continued)

Table 15. (continued)

Solution	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
	Analyses of FG1 Data			Analyses of FG2 Data		
	One-case Threshold, Consistency Threshold of .8			One-case Threshold, Consistency Threshold of .8		
Intermediate	14	$\sim x1*\sim x3$ $\sim x3*x4$ $\sim x1*x2*x4$.8016, .8474	8	$x2*\sim x3*x4$ $x1*x2*\sim x3$ $x1*x2*x4$.6152, .9268
Parsimonious	11	$\sim x3$ $\sim x1*x2$ $\sim x1*x4$ $x2*x4$.9332, .756	12	$x2*\sim x3$ $x1*x2$.6152, .9268
Empirical calibration using FG2 data Complex	10	$x2*x3$ $x1*\sim x3*x4$ $\sim x1*x2$.7645, .9434	11	$x2*x4$ $\sim x1*\sim x2*\sim x3$ $x1*\sim x3*\sim x4$ $\sim x1*x2*x3$.7599, .9085
	10	$x2*x3$ $x1*\sim x3*x4$ $\sim x1*x2$.7645, .9434	10	$x2*x4$ $\sim x1*\sim x2*\sim x3$ $x1*\sim x3*\sim x4$ $\sim x1*x2*x3$.7599, .9085
Parsimonious	10	$\sim x3$ $x2$.9357, .9304	11	$\sim x3$ $x2$.8914, .8983

Note: FG1 = fuzzy group 1; FG2 = fuzzy group 2.

Looking down the FG1 analyses, however, we also find that calibration matters. For example, in the complex solutions, only 3 of the 7 causal recipes appear in two of the three calibrations, and none appear in all three. Matters barely improve when we move to the FG2 data; of the 10 causal recipes in the complex solutions, only 2 match across two calibrations, and again, none match across all three calibrations.

All told, we produce 7 unique recipes for FG1 data, and 10 unique recipes for FG2 data, simply by perturbing the calibration anchors. Thus, complex solutions identify 17 unique recipes or, put another way, depending on the calibration, the outcomes of the 40 cases appear to have been produced through 17 unique causal recipes. In sum, fsQCA appears unable to replicate findings across data sets drawn from the same population.

One might maintain that the problem is that calibrations are based on subsets of the data; if that is the problem, using all of the data in the calibration process should make group-specific analysis findings match. Table 16 contains results based on calibrations based on all the data.

Furthermore, to assess the sensitivity of the results to calibration assumptions, we used two additional calibration alternatives: a complementary log-log distribution calibration of all X 's and a complementary log-log distribution calibration of both the X 's and Y . We thus compare results from calibrations using logit (symmetric) and complementary log-log (asymmetric) distributions.

Results for the logit calibration indicate that using all of the data in the calibration process does not produce stable results across groups. Only one of the nine complex solutions appears in two different analyses, and none appear in all three.

Table 16 also indicates that complementary log-log calibration results do not match logit calibration results. Taken together, these results suggest that one should provide an explicit theory that justifies the specific fsQCA calibration, because results are highly sensitive to calibration method. If one lacks such a theory, any results produced are arguably arbitrary.

9.3. Discussion of Study 4 Results

9.3.1. *The stability of calibration.* Ragin (2000) argues that analysts should use an explicit theory of degrees of inclusion in a set to calibrate measures. Notably, fsQCA recognizes that beyond some threshold,

Table 16. Study 4, Calibration Assessment, Fuzzy Groups 1, 2, and All Data, All Data Empirical Calibration via Logistic and Complementary Log-log, One-case Threshold, Consistency Threshold of .8

Calibration Source/Solution	Proposed (Logit) Calibration			Complementary Log-log Calibration (X)			Complementary Log-log Calibration (X and Y)		
	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
FG1 data Complex	11	$\sim x1*x2$.7449, .8229	4	$x1*x2*x3*x4$.4430, .8419	4	$x1*x2*\sim x3$ $x1*x2*x4$.7263, .9514
		$\sim x1*\sim x3*\sim x4$							
		$x1*\sim x3*x4$							
Parsimonious	11	$x1*\sim x2*x3*\sim x4$.8951, .8128	4	$x3*x4$.4690, .8207	4	$\sim x3$ $x4$.7882, .9164
		$\sim x3$							
		$\sim x1*x2$							
FG2 data Complex	11	$x1*\sim x2*\sim x3$.7642, .8226	No solution possible, consistency threshold not met			4	$x1*x2*\sim x3$ $x1*x2*\sim x4$.6361, .8879
		$\sim x1*\sim x2*x3*\sim x4$							
		$x2*x4$							
Parsimonious	11	$x2*\sim x4$.7762, .7788	No solution possible, consistency threshold not met			4	$\sim x3$ $\sim x4$.6584, .8786
		$x2*\sim x3$							
		$x1*x2$							
		$\sim x1*\sim x3*\sim x4$							
		$\sim x1*\sim x2*\sim x4$							

(continued)

Table 16. (continued)

<i>Calibration Source/Solution</i>	R	Causal Recipes	Coverage, Consistency	R		Coverage, Consistency	
				Proposed (Logit) Calibration	Causal Recipes	Log-log Calibration (X)	Complementary Log-log Calibration (X and Y)
<i>All data</i>							
Complex	15	~x3 x1*x2 x2*x4	.9544, .7669	No solution possible, consistency threshold not met	4	x1*x2*~x3 x1*x2*x4	.7355, .8792
Parsimonious	15	~x3 x1*x2 ~x1*x4 x2*x4	.9544, .7438	No solution possible, consistency threshold not met	4	~x3 x4	.7923, .8668

Note: Italics indicates the data source for the calibration of the variables.

additional change in a variable's value has no impact on whether a case is or is not in the set (Ragin 2008a). Lieberman (2001) agrees but notes that this characteristic can be treated in statistical analyses (e.g., top coding) but is often overlooked. Seawright (2005) notes that variable calibration depends on assumptions equivalent to the functional form assumptions of regression modeling.

Unlike age, marital duration, and other variables whose scale arises from the phenomenon itself, many variables lack an inherent scaling. Researchers have long studied the scaling of such variables (e.g., Clogg 1982; Duncan 1961, 1979; Hauser and Warren 1997; Xie 1992).¹⁷ Calibration is important, and many analysts attend to it.¹⁸

Ragin (2008:87) prefers logit scaling because of symmetry about the 50% crossover point and the translation into unbounded log odds. Yet symmetry and unboundedness are themselves theoretical claims and may often be inappropriate. Asymmetric scales are often more appropriate; for example, if schools are effective, the distribution of student achievement should be asymmetric, for the lower tail should be thinned by schooling. Single- or double-bounded scales can also be preferable; for example, there is an absolute zero to the amount of pollution a nation may produce. Seen in this way, the principles of symmetry and unboundedness used to select the logit scale appear arbitrary. And, if principles of selection are arbitrary, failure to replicate under change of scale renders findings also arbitrary.

One might regard our finding that calibration matters of trivial interest. The importance of the finding, however, stems from the counsel to use the logistic scale for calibration. Any general counsel must be rejected, because for measurement to be truly based in theory, the functional form for measurement calibration must itself be defensible, not based on rote application. Otherwise, sensitivity of findings to calibration implies arbitrariness of findings.

9.3.2. Interaction effects and complex conjunctures. An enduring empirical question is whether the full effect of a given X is visible regardless of the values of Z or, instead, only by jointly considering Z . Such concerns motivate multivariate techniques that control Z to perhaps more clearly reveal the X, Y association. The concern is also manifest in the question of whether one must interact X and Z to understand Y .

Alas, we find much confusion about interactions in the QCA literature. For example, Vaisey (2007) claims a major difference between

interactions and complex conjunctural conditions. Although their imagery may differ, they cannot be distinguished empirically, for a conjunctural condition is one combination of an interaction.

Vaisey (2007:862, note 11) also contends that the $X_1 \times X_2$ interaction has the same values for high X_1 or high X_2 , preventing one from distinguishing different effects for high X_1 versus high X_2 . Although this claim is true under many common codings of X_1 and X_2 , the claim more fundamentally ignores the flexibility of modeling. Given theoretical warrant, one might estimate the $X_1 \times X_2$ interaction plus an $X_1 \times X_3$ interaction, defining $X_3 = 1$ when $X_1 > \omega$ and zero otherwise, where ω references the cutoff beyond which one theorizes a change in the relationship. With this specification, one could distinguish between an interaction effect when X_1 is high and when X_2 is high. Because $X_{k \geq 3}$ could be defined otherwise, this is only one of many possibilities available.

Amid such confusion, what must not be lost is the finding that QCA routinely selects interactions when outcomes were produced with main effects only. We suggest that whatever causes QCA's routine selection of interactions may contribute to the instability of QCA results.

9.3.3. Replication and case-oriented research. Some claim that case-oriented researchers are unconcerned with replication because they are intrinsically interested in the cases they study. Yet QCA's failure to replicate results means that the cases, plural, in the sample—not the causal structure of any specific case of interest—drive the results. This would seem to pose a problem for case-oriented researchers.

QCA analysts might counter by contending that the very sensitivity of analysis results to the set of cases considered explains why analysts prefer purposive to probability sampling. Purposive sampling, however, does not resolve the issue, for its use means that if analysts have different purposes (e.g., assessing whether theory X explains Y vs. assessing whether theory Z explains Y) and use different cases from the same population (e.g., leading economies in 2009), QCA may produce different, perhaps contradictory findings. Even though the studies address the same population (e.g., leading economies) and issue (e.g., austerity conflict), purposive sampling offers no assurance that results will converge to one story, even if one story applies to both sets of cases. Case-oriented analysts could reject the population concept as inappropriate, but doing so makes comparison questionable, for if one conducts comparative analyses, one must identify a set of comparable cases, even if

one uses a synonym (e.g., collection or set) for the concept. It is this interest in comparability, definitive of all comparative research, that makes replication important, as comparable sets of cases should render equivalent results.

Thus, from a case-oriented perspective, that QCA does not replicate the known causal process for cases in equivalent sets means that findings for a case of interest depend on the other cases analyzed. This should pose a major problem for case-oriented studies, for it means that QCA findings for, say, the French Revolution, differ depending on whether, for example, the 1917 Russian Revolution is in or out of the sample. Thus, the replication failures indicate that QCA may totally mischaracterize the causal process of studied cases.

10. STUDY 5: REPLICABILITY, STABILITY, MEASUREMENT ERROR, AND EFFICACY IN CRISP-SET QUALITATIVE COMPARATIVE ANALYSIS

We reconsider replicability and stability because fsQCA is so sensitive to calibration that its failure to replicate could reflect a calibration failure on our part. Because csQCA does not require calibration, the focus is not blurred. We further address measurement error.

10.1. Study 5 Methods

We constructed a 40-case data set using a known set of causal relations and randomly allocate the 40 cases to crisp group 1 (CG1) or crisp group 2 (CG2). Cases in CG1 and CG2 were produced via the same causal process. (The CG2 data were used in study 2.) For study 5, each author separately analyzed one of the two data sets. Afterward, we reanalyze the combined data to give QCA another chance to find the correct causal recipes. Finally, we explore the stability of results using delete-one-at-a-time (D1) simulation, a strategy akin to jack-knifing.

10.2. Study 5 Results

Results for CG1 data are in Table 17, quadrant I. Each solution has three causal recipes. Complex and intermediate solutions are identical, differing slightly from the parsimonious one. Quadrant IV contains results for the CG2 data. Because the cases were randomly assigned to CG1 and

Table 17. Study 5, Results for Two Random Samples of 20 from One Simulated Data Set of $n = 40$

Solution	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
CG1 data						
		Quadrant I method 8			Quadrant II method 6	
		First analysis of group 1 data: one-case threshold, consistency threshold of .8			Second analysis of group 1 data: one-case threshold, consistency threshold of .6	
Complex	11	$\sim x2 * \sim x3 * \sim x4$ $\sim x1 * x2 * \sim x4$ $x2 * \sim x3 * x4$.7692, 1.0000	11	$\sim x3 * \sim x4$ $x2 * \sim x3$ $\sim x1 * x2 * \sim x4$.9231, .9231
Intermediate	6	$\sim x2 * \sim x3 * \sim x4$ $\sim x1 * x2 * \sim x4$ $x2 * \sim x3 * x4$.7692, 1.0000	10	$\sim x3 * \sim x4$ $x2 * \sim x3$ $\sim x1 * x2 * \sim x4$.9231, .9231
Parsimonious	11	$\sim x1 * \sim x4$ $\sim x2 * \sim x3 * \sim x4$ $x2 * \sim x3 * x4$.7692, 1.0000	11	$\sim x3 * \sim x4$ $\sim x1 * \sim x4$ $x2 * \sim x3$.9231, .9231
CG2 data						
		Quadrant III method 8			Quadrant IV method 6	
		Second analysis of group 2 data: one-case threshold, consistency threshold of .8			First analysis of group 2 data: one-case threshold, consistency threshold of .6	
Complex	13	$x2 * \sim x4$ $\sim x1 * x2$ $x1 * \sim x3 * \sim x4$.5385, 1.0000	13	$\sim x3 * \sim x4$ $\sim x1 * x2$ $x2 * x3$.9231, .8571
Intermediate	10	$\sim x1 * x2$ $x2 * \sim x4$ $x1 * \sim x3 * \sim x4$.5385, 1.0000	12	$\sim x3 * \sim x4$ $\sim x1 * x2$ $x2 * x3$.9231, .8571
Parsimonious	13	$x1 * \sim x3$ $x2 * \sim x4$ $\sim x1 * x2$.5385, 1.0000	13	$x2$ $\sim x3 * \sim x4$.9231, .8571

Note: CG1 = crisp group 1; CG2 = crisp group 2.

CG2, the analysis of CG2 data should obtain the same causal recipes that were produced by analysis of cases randomly assigned to CG1. However, none of the causal recipes in quadrant IV for CG2 match the causal recipes found in quadrant I for CG1 data.

Quadrant IV is the result of different analysts' decision making while unaware of the findings and procedures for the analysis of CG1. The only analytic decision on which the researchers diverged is that the CG2 analyst chose a consistency threshold of .6 rather than .8. One concern is whether this lone specification difference underlies the differences between quadrants I and IV. The decision is potentially consequential; in the fsQCA manual, Ragin (2008b:46) indicates that consistency scores below .75 reflect nontrivial inconsistency. The analyst of CG2 data accepted a lower consistency threshold because one configuration of the data had a consistency score of .667, just below the .75 cutoff, while the next highest consistency score was .333, substantially lower. Still, because this decision might have so damaged the enterprise that it undermined replication, we conducted additional analyses.¹⁹

In Table 17, quadrants II and III reveal what each analyst found when using the analytic decisions that had been made in the analysis of the other group. If the different analytic decisions drive the differences between quadrants I and IV, then the solutions for quadrants I and III should match each other, and the solutions for quadrants II and IV should match each other.

Considering quadrants I and III first, for their results are based on the more stringent, more advised consistency threshold, we find that not one of the four causal recipes appears in both quadrants. Quadrants II and IV are also illuminating. One of the four recipes obtained in quadrant IV is found in quadrant II, but the other two recipes in quadrant IV are not in quadrant II, and two recipes in quadrant II are not found in quadrant IV. It seems the widely divergent results are well matched to their samples, just as one would expect were overfitting occurring.

The analyses produced 12 different causal recipes, of which only 1 ($X_3 = X_4 = 0$) is replicated across the two data sets. Even so, none of the causal recipes match the known causal structure. Collecting the results, QCA fails to replicate causal recipes across data sets.

Combining CG1 and CG2 data allows us to directly determine whether the findings above are a result of problems produced by randomly allocating cases to groups. This is an important hypothesis to assess, for some analysts view QCA as designed for case-oriented

research. Many such analysts reject probability sampling, rendering any analysis using random allocation potentially less persuasive. Moreover, the subsamples were necessarily small, lowering their precision. Finally, many QCA studies may use the full population, and it could be that with the full population, the issues above dissipate. Thus, we reassess the matter using all 40 cases.

Table 18 presents the complex and parsimonious solutions for two analyses of the full population. Considering the one-case threshold analysis first, three of the five recipes do not appear in the earlier group-specific analyses. All of the causal recipes involve at least three factors; one of the new recipes that does not appear in the subsample analysis involves all four factors, which suggests that adding cases may cause the algorithm to identify more complex causal recipes.

The two-case threshold analysis produces only three rather than the five recipes identified in the former analysis. Two of the three recipes are not observed in the subsample analyses. Again, a four-way interaction is a chosen causal recipe.

To further consider the stability of the estimates, we conducted a D1 analysis. This approach resembles a Monte Carlo analysis in that results are obtained for multiple subsets of the data, but it is not a Monte Carlo analysis in that we obtain all results possible under a given (and in this analysis a one-case) deletion regimen. Monte Carlo simulations may use only a small proportion of the possible subsets of the data, relying instead on statistical theory to draw inferences. Here, however, relying on the same statistical inference logic QCA analysts may question might only increase the distance between our analysis and QCA proponents' claims and thus seems inappropriate. Consequently, we opt for D1 simulation instead.

Considering the complex solution (see Table 19), of the 81 causal recipes possible, we consistently obtained the same 5. Fully 95 percent of the replicates produced solutions 1, 2, 3, 5, and 6. The parsimonious solutions were similarly concentrated, identifying the same four causal recipes 95 percent of the time. Thus, QCA D1 results are much more stable, indicating that these recipes accurately reflect QCA analysis of the data.

One solution that appears in 95 percent of the complex solutions is a four-way interaction, and all of the other "95 percent" solutions of the D1A analyses involve three-way interactions. Indeed, all solutions identified in study 5, except for one, involved conjunctions of two or more variables. Yet the actual causal structure producing $Y = 1$ set coefficients

Table 18. Study 5 Analysis of Crisp Simulated Data Set of $n = 40$ (All Sample)

Solution	One-case Threshold, Consistency Threshold of .8			Two-case Threshold, Consistency Threshold of .8		
	R	Causal Recipes	Coverage, Consistency	R	Causal Recipes	Coverage, Consistency
Complex	14	$x2 * x3 * \sim x4$.5385, 1.000	11	$\sim x1 * x2 * \sim x4$.4615, 1.000
		$x2 * \sim x3 * x4$			$\sim x1 * x2 * \sim x3$	
		$x1 * \sim x2 * \sim x3 * \sim x4$			$x1 * \sim x2 * \sim x3 * \sim x4$	
Parsimonious	14	$\sim x1 * x2 * \sim x4$.5385, 1.000	11	$x1 * \sim x2 * \sim x3$.4615, 1.000
		$\sim x1 * x2 * \sim x3$			$\sim x1 * x2 * \sim x4$	
		$x1 * \sim x2 * \sim x3$			$\sim x1 * x2 * \sim x3$	
		$x2 * x3 * \sim x4$				
		$x2 * \sim x3 * x4$				

Table 19. Study 5, Delete One Analysis, Crisp All Sample, One-case Threshold, Consistency Threshold of .8

	Ingredients and Recipes				Count of Solutions	
	X_1	X_2	X_3	X_4	Complex Result	Parsimonious Result
1	—	1	1	0	38	38
2	—	1	0	1	38	38
3	1	0	0	0	38	0
4	1	0	0	—	0	38
5	0	1	0	—	38	38
6	0	1	—	0	38	1
7	1	—	0	0	1	0
8	—	1	0	—	1	1
9	—	1	—	0	1	1
10	0	—	0	1	1	0
11	—	0	0	0	1	1
12	1	—	0	—	0	1
13	0	—	—	0	0	1

Note: In the recipe, if $X = 0$, it is scored 0; if $X = 1$, it is scored 1; if $X = \text{—}$, its value is irrelevant to the recipe.

for all interactions equal to zero. Thus, results are usually unstable, yet even stable results are consistently incorrect.

10.3. *Discussion of Study 5 Results*

10.3.1. *Frequency of cases thresholds and measurement error.* Measurement error is an issue for all research methods. For example, measurement error in linear regression models attenuates coefficients and renders them inconsistent.²⁰

Critics contend that measurement error is especially debilitating for QCA because imperfect measurement could cause a case to be misclassified. The higher the consistency threshold, the more easily a misclassified case can distort findings. One partial response to this problem is to remove causal configurations that fall beneath a frequency of cases threshold. Yet this strategy is ineffective.

In Figure 1, numbers signify the errorless location of a nation on X^* , a latent variable mapped onto the dichotomous variable X . Dots indicate the location an observer believes the nation falls on the latent variable; in fsQCA terms, this could be the set membership score. Values to the right of the vertical line fall inside the crisp set of X , while values to the

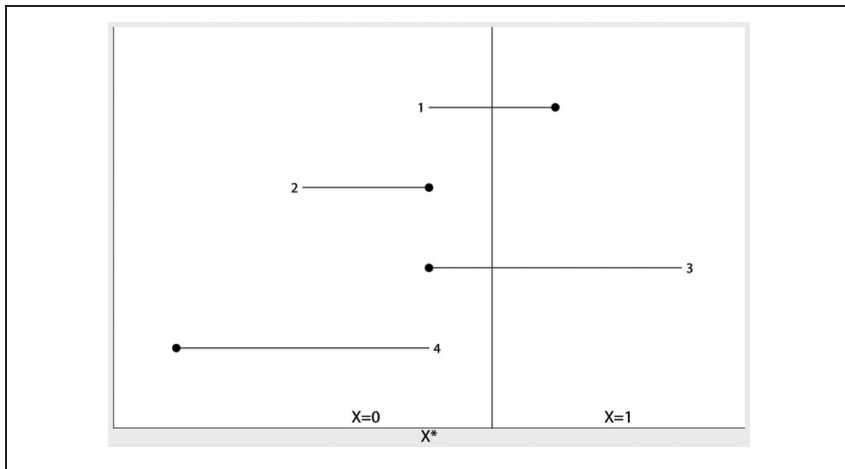


Figure 1. Representation of error for four nations.

left of the vertical line fall outside the crisp set of X . Error is always potentially consequential for fsQCA but is consequential for csQCA only if it flips X from its actual value to its opposite. Focusing on csQCA, two conditions increase the chance of a flipped value: (1) the errorless measure is close to the threshold or (2) the error is large. In Figure 1, nation 1 has a small error but because the nation is close to the threshold error moves its observation from $X=0$ to $X=1$. Nation 2 has the same amount of error, but it is inconsequential because nation 2 is far from the threshold. Nations 3 and 4 illustrate the same possibility for large errors.

For csQCA, errors can be large and have no effect or small and have big effects. This is always true for any threshold method. Because all cases have both informational content and error, Figure 1 also implies that removing rare causal configurations is insufficient, for it leaves intact the error in the cases in the remaining causal configurations.

QCA measures configuration consistency with the conditional mean of Y . By the central limit theorem, the more cases in the configuration, the more precise its estimated conditional mean. Thus, the implicit assumption behind deleting rare causal configurations is that error may play a bigger role in their consistency scores than in the consistency score for common configurations. Hence, removing rare causal configurations removes configurations most likely to harm QCA analyses because of measurement error.

Yet problems ensue. First, there is no necessary connection at the case level between size or consequentiality of error on one hand and placement in a rare configuration on the other. Second, deleting rare configurations removes configurations that might be most crucial for understanding the process at issue, because a case in the rare configuration is arguably far more informative than is yet another case in the common configurations. Thus, the deletion strategy throws away information, is poorly targeted, and may reduce theoretical leverage. Consequently, measurement error harms QCA, and one response to the problem further harms QCA.

10.3.2. Replication. QCA failed to distinguish significant combinations from spurious patterns. Solutions were exceedingly volatile, although we find far less volatility in the D1 analysis. Still, the D1 analysis revealed that QCA repeatedly selected the same erroneous set of causal recipes containing multiple three-way interactions and one four-way interaction. These results raise major questions about QCA, including Lieberman's (2004) question: Will QCA find causal recipes for non-causal data? It is to that question we now turn.

11. STUDY 6: SEARCHING FOR CAUSAL PATTERNS IN NONCAUSAL DATA

Both determinists and stochasticists view causal relations as systematic relations. Although determinists' requirement for a perfect association of cause and outcome implies that causal relations are systematic relations, stochasticists follow a different path to the same conclusion. Stochasticists posit first that factors omitted from an explanation are jointly random and thus ignorable. Yet they also allow separate omitted factors to perhaps have patterned relations with outcomes. If plots or other information reveal a systematic relation between an omitted factor and the outcome (or unexplained aspects of the outcome), the formerly omitted factor may be added to the model or explanation. This modeling procedure indicates that although stochasticists do not require perfect correlation of cause and outcome, they still require systematicity. Hence, both determinists and stochasticists require systematicity for causality to pertain.

At the same time, researchers are not infallible; they may believe that a noncausal factor is causal. Thus, if researchers are not always and everywhere able to ensure their variables are causal, methods must be

Table 20. Study 6, Causal Recipes and Summary Statistics for Analysis of Noncausal Data, One-case Threshold, Consistency Threshold of .8

Solution	Rows	Causal Recipes	Coverage	Consistency
Complex	7	$x_1 \sim x_2 \sim x_3$ $\sim x_1 \sim x_2 \sim x_3$.4444	1.0000
Intermediate	2	$\sim x_3 \sim x_2 \sim x_1$ $\sim x_1 \sim x_2 \sim x_3$.4444	1.0000
Parsimonious	7	$x_1 \sim x_3$ $\sim x_1 \sim x_2 \sim x_3$.4444	1.0000

able to separate systematic from nonsystematic relations, as all agree that only the former can be causal.

Juxtaposing the agreement that systematicity is required for causality with human fallibility raises an important question: Will QCA be able to show that a specification is bereft of causal factors? Or, put starkly, will QCA find causal recipes in data so lacking in causal relations as to appear random?

11.1. Study 6 Methods

We generated a noncausal data set (NCD), such that Y has no causal relationship to the X 's, by randomly assigning cases to 0 or 1 on each variable. One should expect the association among the variables to be zero, but, randomly it might not be. Yet, correlations and partial correlations confirm that X_1 , X_2 , and X_3 do not have a statistically significant linear association with Y . Because all variables are dichotomous, there is insufficient measurement precision for the variables to reflect a curvilinear relationship.

11.2. Study 6 Results

Complex and intermediate solutions are identical—"X₁ and not X₂ and not X₃" and "not X₁ and X₂ and X₃,"—and have an overall coverage of .44 and a consistency of 1.00 (see Table 20).²¹ The solutions are inverses of each other. The parsimonious solution is slightly different: "X₁ and not X₃" and "not X₁ and X₂ and X₃." Thus, in the analysis of noncausal data, QCA has generated multiple causal recipes that reflect the technique's ability to detect conjunctural patterns correlational analysis would not. However, in this case, the patterns are spurious; the

method detects complex causality even though the data are noncausal, which is the outcome Lieberman anticipated. The results raise serious doubts about QCA's ability to avoid certifying spurious causal patterns.

11.3. Discussion of Study 6 Results

11.3.1. *The roll of the dice? The French Revolution? Spurious causation?* We revisit determinism, stochasticism, and QCA in the context of considerations concerning spurious causation. Mahoney (2008:416–17) contends that the actual probability of an event that occurs is 1. We learn that $\Pr(Y = 1) = 1$ only after the event occurs, but at the time of the event, $\Pr(Y = 1)$ was 1. So, for example, if a gambler rolls a 2 and a 3 at 10:04 p.m. Pacific time on Thursday at a craps table in Las Vegas, we know that the probability of that outcome at that table and time was 1. The way the gambler held the dice, the strength of the throw, and more coalesce such that “at the time of the release of the dice” (Mahoney 2008:416), the specific case outcome that occurred was a certainty. Similarly, the probability of the French Revolution of 1789 was, at some moment prior to its occurrence, equal to 1; after the revolution, we learn that it was a certainty. Mahoney (2008) thus contends that at the case level, at least, chance does not exist, such that case-oriented research requires a deterministic approach. We call the moment at which Mahoney claims the outcome (O) is completely determined P, for privileged.

There are many problems with Mahoney's formulation. For one, positing P accounts for everything between P and O, partly addressing determinists' need to explicitly treat everything prior to O, though everything before P still requires attention. Yet positing P means that even an earthquake measuring 7.0 on the Richter scale that occurs half a second after P cannot affect the final orientation of the dice, an implausible proposition. As another example, chance is allegedly nonexistent after P, but what about before? Was chance able to cause a different dice roll before P? If not, why study causes, as all outcomes are predestined, yet, if so, what kind of determinism posits chance before some magic moment, fatedness after? Given such conundrums, we contend that although determinism is conceivable, inconsistencies haunt serious efforts to engage it. Strategies exist to address the inconsistencies; space constraints preclude our describing them and conveying why we contend that they fail.

However, even if one accepts the claim that chance is nonexistent at the case level, we know of no one who claims that the nonexistence of

chance means that every prior factor in the universe is causal for the dice roll at a Las Vegas craps table or for the 1789 French Revolution. Indeed, the positing of P reflects determinist's affirmation of this broad consensus: some preexisting *X*'s are noncausal for some *Y*'s. Thus, despite the entertaining episodes of the old television series *Connections* (Burke 1978), Ben Franklin's dog scratching its left ear at 11:07 a.m. on July 12, 1788, is (probably) not causal for the decision to publish the first Harry Potter book in 1997, nor for ratification of the U.S. Constitution in 1789.

Yet, given human fallibility and determinists' denial of chance, epistemological determinists can only maintain that QCA need not eliminate noncausal factors by rejecting P and positing pancausality; that is, everything causes everything. These two moves work because if one both rejects P and adopts pancausality, one may claim that QCA need not eliminate noncausal factors because there are no noncausal factors. Yet accepting pancausality implies that QCA, the general linear model, and all divisions of phenomena into cause and effect are at best misleading and should be replaced by neocybernetics (e.g., Hyötyniemi 2006).

Pancausality is an assumption one must make to render irrelevant the results of study 6. If one rejects pancausality, then it is possible that a researcher may believe that a factor is causal for a phenomenon but be incorrect in that belief. Given this possibility, analysts' methods must be able to distinguish systematic from nonsystematic connections, for only the former can be certified as causal. Some analysts regard some nonsystematic connections as mere chance, but even if one is a determinist and denies the category of chance connections, as long as incorrect causal claims are possible, our methods must be able to distinguish systematic and nonsystematic connections. However, study 6 indicates that QCA cannot make this distinction.

12. AN EMERGING FUNDAMENTAL ISSUE: IS QUALITATIVE COMPARATIVE ANALYSIS TRULY CASE ORIENTED?

Many claims about QCA are articulated in a context sculpted, in part, by contention that the method is designed for case-oriented analyses. Achen (2005) contends that case-oriented research is often extremely illuminating and powerfully persuasive. We agree.

Case-oriented researchers aim to understand the specific case(s) analyzed (e.g., the French Revolution, the 10:04 p.m. roll of the dice), not other cases (e.g., revolutions in general, other dice rolls). Case-oriented researchers often select cases purposively, not through probability sampling. Although purposive case selection is defensible for case-oriented research, we have already shown that such sampling clashes with comparative logic (see pp. 51–52).²²

Seeing QCA as case oriented, which justifies its use on nonprobability samples and its neglect of replication, is, we submit, a major source of its appeal, for there are few systematic means for analyzing data in a case-oriented framework. Were QCA to correctly identify causal recipes or assess asymmetric causation, as a standardized case-oriented method, it would be a breakthrough, perhaps even a step toward unifying social scientists' methodological dialogue.

Yet we find no evidence that QCA is case oriented. Its proponents argue that QCA is case oriented because analysts are advised to return to the specific cases during their research (Ragin 2008a:21; Schneider and Wagemann 2010; Vaisey 2007). This is a weak basis for the claim, however, for QCA analysts do not always return to cases (e.g., Ragin 2006, 2008a; Dy et al. 2005). Furthermore, any analyst can use most any method and then, at some point in the broader inquiry, peruse specific cases. The counsel for conducting OLS regression analysis says as much. Neter, Wasserman, and Kutner (1989) wrote,

Outliers are extreme observations. . . . When we encounter one, our first suspicion is that the observation resulted from a mistake or other extraneous effect, and hence should be discarded. . . . On the other hand, outliers may convey significant information, as when an outlier occurs because of an interaction with another independent variable omitted from the model. A safe rule frequently suggested is to discard an outlier only if there is direct evidence that it represents an error in recording, a miscalculation, a malfunctioning of equipment, or other similar type of circumstance. (pp. 121–22)

The only way to assess whether such recording malfunctions occurred is to delve deeply into the case. Furthermore, to assess whether interactions or other omitted variables help explain the case one must incorporate additional factors into the model, and this too may necessitate exploration of various cases. If looking more closely at one or more

cases is sufficient to view the operation of a method as case oriented, then OLS regression modeling is a case-oriented method.

We do not believe this is what case-oriented researchers really mean when they say that their work is case oriented, and thus we regard neither OLS regression nor QCA as case oriented.

The non-case-oriented nature of QCA is latent, but we can excavate it by attending closely to the advice QCA analysts receive and to the actual operation of QCA. One example is visible in Ragin's (2008a) advice concerning the frequency of cases threshold:

The key task in this phase of the analysis is to establish a number-of-cases threshold, that is, to develop a rule for classifying some combinations of conditions as relevant. . . . The rule established by the investigator must reflect the nature of the evidence and the character of the study. Important considerations include the total number of cases, the number of causal conditions, the degree of familiarity of the researcher with each case, the degree of precision that is possible in the calibration of fuzzy sets, the extent of measurement and assignment error, whether the researcher is interested in coarse versus fine-grained patterns in the results, and so on. (pp. 131–32)

Ragin (2008a) further notes that when one has few cases for investigation, one should use a frequency of cases threshold of "1 or 2" (p. 143).

Yet what if the frequency threshold throws France out of a study of revolutions or the United States out of a study of dominant world powers? Would a true case-oriented method apply a frequency criterion to the construction or selection of its explanandum? How can a method be case oriented if it can prevent researchers from studying cases of intrinsic interest if they lack *other* similar cases in the data set? Although the counsel suggests that one consider many factors in setting thresholds, nowhere in the list of factors is the explicit claim "Do not set thresholds that delete focal cases." Can a method be case oriented if its advised use abets deleting the very cases one seeks to study?

As another example, QCA analysts have added statistical testing to the capability of QCA (Longest and Vaisey 2008; Ragin 2000:107–18). But what is the point of a statistical test of whether peasant uprising caused a discernibly different from zero number of the French, Russian, Chinese, American, and Cuban revolutions if interest lies in those specific cases? What relevant question does statistical test of QCA results answer here? Adding statistical test capability may make QCA seem

familiar to many, but statistical testing has little place in case-oriented studies.

These are but two of several indications that QCA is not really case oriented. QCA is, instead, configuration oriented, that is, cell oriented. After all, it is the configuration—a cell of a cross-classificatory array—for which one implements the frequency threshold, the configuration for which one specifies the consistency threshold, and the configuration that is translated into causal recipes. It is the configuration one decides to exclude or include, and cases are along for the ride—or ancillary removed before the ride when their cells are removed, as our discussion of measurement error noted. And, as configurations are defined completely by variable interactions, QCA is variable oriented or, at best, interaction variable oriented. Thus, QCA appears no more case oriented than latent class analysis or log-linear modeling. The cell of a cross-classification is the fundamental unit of analysis for the latter two methods, and it is the fundamental unit of analysis for QCA as well. To maintain otherwise QCA analysts must show the use of cases within the method, distinct from the use of causal configurations, as causal configurations are but zones in a multidimensional space defined by multiple factors on which the cases are measured.

These observations distinguish QCA and case-oriented research. Yet QCA proponents sometimes claim that QCA combines qualitative and quantitative methods (e.g., Ragin 2008a) and thus may dismiss our observations by claiming the aspects we identify (e.g., frequency threshold advice) reflect the quantitative side of QCA. If so, we wonder, what is QCA's qualitative side? Evidence suggests that if QCA really is a synthesis of qualitative and quantitative methods, it is still one from which cases recede from view even as important comparisons are ignored, violating the principles and interests of both qualitative and quantitative research at once. Thus, even seen as a synthesis, QCA fumbles away the advantages of both methodological traditions.

Hence, we contend that it is imperative that one distinguish QCA on the one hand and case-oriented methods on the other. Once distinguished, it is clear that one may reject QCA and embrace case-oriented research. Indeed, our view is that the former act is essential to the latter.

Once the cell-oriented nature of QCA is recognized, there is no warrant for its use on nonprobability samples, nor can its use be justified by claiming that results apply only to the cases studied. These conclusions dissolve any remaining justification for QCA.

13. CONCLUSION: EPISTEMOLOGICAL BAIT AND METHODOLOGICAL SWITCH

Critics and proponents take QCA's claimed epistemological commitments as accurately reflecting the method and proceed to debate the wisdom of those commitments. Closer inspection reveals that QCA fails to follow through on multiple claimed epistemological commitments.

QCA is sometimes described as stochastic (e.g., Ragin 2000:109), but is sometimes described as deterministic (e.g., Mahoney 2008; Ragin 2008a:44). Vacillation ensures disjuncture. However, as shown above, whatever the claim, QCA findings provide no evidence of deterministic cause, because QCA relaxes the severe demands deterministic methods necessarily entail.

QCA is also described as case oriented. Yet calling QCA case oriented is like calling log-linear modeling case oriented. No matter how much knowledge of specific cases an analyst brings to the investigation, QCA operations fundamentally handle only the cells of the table, not the individual cases inside the cells. Thus, claims of case orientation mischaracterize QCA.

The method's appellation, qualitative comparative analysis, suggests a focus on key comparisons. In fact, the method systematically impedes assessment of the central comparisons causal investigations require by, for example, failing to compare outcomes for configurations that analysts specify as causal to outcomes associated with configurations analysts doubt are causal. Lacking this comparison and other essential ones, the method is deeply noncomparative.

Finally, QCA claims to excavate causal asymmetries. Most uses of the method use only cross-sectional data, so dynamic asymmetry is inaccessible, and the method's approach—study the determinants of $Y = 1$ and $Y = 0$ (or $Y = \text{high}$ and $Y = \text{low}$) separately—leaves functional form asymmetry inaccessible as well. What is left is an illusory asymmetry produced by specification of the consistency thresholds.

Thus, the QCA claim to be a deterministic/stochastic, case-oriented, comparative analytic tool for exploring asymmetric causality provides the bait. QCA is actually a self-contradictory, cell-oriented, noncomparative, nonanalytic means to identify asymmetric causal illusions—this is the switch. The bait brings proponents and critics into debate, but the switch reduces the ability of their debate to illuminate both the research operations of QCA and the viability of divergent

epistemological positions. The epistemological debate is worthwhile, for the issues in play are complex, and only through their extended examination may their tractability, costs, and benefits become clear. However, the bait and switch, by unhinging the method from its rhetorical moorings, damages understanding and hinders methodological development.

The bait and switch is also substantively consequential, for many researchers rely on proponents' descriptions of QCA. Because of those descriptions, analysts have used csQCA and fsQCA to study diverse issues, including organizational influences on ergonomic injuries (Marx and van Hootegem 2007), social policy development (Amenta and Halfmann 2000); conditions of HIV/AIDS incidence in Africa (Cronqvist and Berg-Schlosser 2006), the development of children's rights support (Gran and Aliberti 2003), and more (e.g., Glaesser et al. 2009; Vaisey 2007). Indeed, the number of QCA analyses appears to be increasing. Yet our pre-*Challenger* launch data reanalysis suggests that QCA studies likely reach faulty conclusions, and our simulations indicate major reasons why. If authorities act on QCA findings, the consequences may be harmful or, in some substantive domains, perhaps even fatal. And correcting the errors may be difficult for, as QCA studies proliferate, solid case-oriented research may be crowded out.

QCA can be interpreted as an attempt to formalize qualitative research; formalization is evident, for example, in QCA counsel for a minimum consistency threshold of .75. Although the gridded data of QCA remains arguably qualitative, the form is that of statistical analysis. As form necessarily constrains, one must always ask whether a proposed form facilitates or impedes research. Formalization is sometimes harmful (Lucas 2012), and that appears to be the case here.

QCA proponents sometimes claim that QCA combines qualitative and quantitative methods. If so, QCA offers a particularly inauspicious combination of the two, purging the strengths and retaining the weaknesses of both approaches, all while creating new, debilitating methodological errors for empirical research.

The dimensions of QCA failure extend beyond the major ones summarized. Of the 70 solutions presented above, QCA rendered the correct causal story in only 3. Two of those cases were found to be artifacts of deficient means of identifying asymmetric causality, and the third required one to know the correct causal story before invoking the

method. Thus, far from successes, these correct solutions show further severe weaknesses of QCA.

Some might dismiss our studies as mere simulations, yet each was designed to study issues of enduring methodological interest, challenges common to real-life research, and/or specific justifications proponents offer for QCA. Thus, study 1 causes are deterministic and sufficient, matching QCA proponents' ontological claims. Study 2 data have symmetric causes, matching the presumption of most published sociology, to discern whether QCA finds spurious asymmetric causation. Study 3 deleted overdetermined cases from study 1 data and in doing so exhumed QCA's hidden, debilitating, implausible "no overdetermination" assumption. Study 4 investigated whether cases produced with the same causal process, but analyzed in subsets using fsQCA, would identify the causal process. Because of the woeful performance of fsQCA, study 5 readdressed replication using csQCA; csQCA also failed. Study 6 studied whether QCA would find causal patterns in non-causal data, to assess how QCA handles a threat to which every analyst is always at risk: mistakenly seeing causality in noncausal conditions. We even included an analysis of real-world data, as QCA proponents advise. Because the three occasions in which QCA obtained the correct solutions revealed serious identification problems or required prior, correct, and complete knowledge of the causal story, even these "successes" actually revealed additional, deeply entrenched reasons for QCA's consistent failures. Thus, we conclude that, methodologically, in every single test QCA failed.

The number of possible combinations of data, functional form, and epistemology is vast. One could maintain that QCA failed because none of our studies hit just the right combination for which QCA is perfectly aligned. We believe that alignment concerns can be taken so far (e.g., deterministic, nonparametric, INUS causation data with moderate R^2 values and known lack of overdetermination) that if required for QCA to work offer yet another critique of the method, suggesting that one must know the answer to one's research question before one can use the method to obtain the answer to one's research question. Although in response to possible alignment concerns we conducted several additional studies (available online), still, some issues are fundamental. No study can address every conceivable permutation; thus, we address fundamental issues. For example, one fundamental issue concerns how a method works under conditions proponents deem ideal. Study 1 found

that QCA fails under such conditions. Another fundamental example asks whether a method spuriously finds causal factors in noncausal data. Study 6 found that QCA failed at this task. Methods that bungle such basic issues impede our understanding and should be rejected.

Of course, analysts bring a wealth of knowledge and imagination to their research such that when results are produced, no matter how convoluted, most analysts can discern a kernel from which to devise a narrative that makes sense of the result. We are reminded of Lazarsfeld's (1949) famous review of *The American Soldier* in which he conveys findings from Stouffer's (1949) study of World War II U.S. soldiers, after which the reader, having had time to accept the findings as obvious, is then informed that every "finding" in the earlier part of the review is the opposite of what Stouffer actually found. We submit that analysts, believing QCA valid on the basis of the claims in its foundational texts, may have inadvertently engaged in similar reconciliation, wrenching coherence from the cacophony of QCA results. However, only reanalyses using other methods will discern which QCA findings, if any, can be sustained.

Multiple better methods, such as narrative strategies (Mahoney 1999) and process-tracing (George and Bennett 2005), are available for comparative research. Analysts routinely use such methods to develop deep sociological insights and to convey their findings convincingly. The obvious case is provided by the narrative moments of Skocpol (1979), but many other examples exist (e.g., Vaughan 1996; Perrow 1999; Adams 2005; Karabel 2005; Mann 2005; Steinmetz 2007; Somers 2008). Any list of such works must be incomplete, but the existence of powerful analyses of macro-causal, historical, and other complex social phenomena is indisputable.

Still, the prospect of formalization raises hope in some quarters that case-oriented research can be systematized via an algebraic system. The hope is that formalization will not obliterate the nuance that is one of the major advantages of qualitative work, that case-oriented scholars, empowered by the formal methods, will apprehend new vistas beyond "general linear reality" (Abbott 1988) as they continue to draw on their detailed knowledge of specific cases. The hope is nurtured by an understanding that many intriguing and important substantive and theoretical questions sorely need such methods.

Yet need is no guarantee that the need will be satisfied, nor that a particular proposal will serve satisfactorily. In this case, we find that the

need is not satisfied by the offered solution. Further, the method is grounded in several dubious epistemological propositions that fail even when ontologically accurate. Thus, we conclude that, despite the hope it represented, researchers should reject QCA and abandon the chimerical quest for a deterministic method.

Acknowledgments

We thank David Grusky for providing data for supplementary analyses and three anonymous reviewers and the editor for helpful comments. An earlier draft of this article was presented at the World Congress of the International Sociological Association, in Session 11 of Research Committee Number 28, Gothenburg, Sweden, July 14, 2010. All errors and omissions are the fault of the authors.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We appreciate the generous support provided to both authors by the American Sociological Association's Travel Award Grant (SES-0852624) supported by the Sociology Program of the National Science Foundation.

Notes

1. Software for QCA includes Ragin's fsQCA, Longest and Vaisey's "fuzzy" Stata ado, and Duşa's (2010) Package "QCA" for R. All use the Quine-McCluskey algorithm, which, if correctly implemented, means that results should not change across programs. Thus, we do not repeat our work on multiple software packages, because doing so would test software, not QCA's utility *per se*.
2. Discovery and justification contexts differ (Reichenbach 1938) but entwine (Gigerenzer 1991). Earlier drafts reflected our discovery context motivation to teach, use, and extend QCA. Yet some questioned our motives. Lacking space to elaborate, we now omit these influences.
3. One reviewer claimed it essential to treat Ragin's (1992) distinction between "casing," the processes researchers use to construct their cases, and cases. Yet, that analysts construct their study objects has been long known (e.g., Thompson and Walker 1982; Laumann, Marsden, and Prensky 1983), such that casing is not distinctive of any methodological tradition. Furthermore, casing is irrelevant here because the input for QCA is information on cases, however constructed.
4. Because fuzzy-set membership can be less than both membership in the outcome and than membership in nonoccurrence of the outcome (Ragin 2008a:137), sometimes fsQCA may find a case to be both a subset of an outcome ($Y \approx 1$) and a subset of a nonoutcome ($Y \approx 0$).
5. Analysts have elaborated a counterfactual framework (e.g., Holland 1986; Morgan and Winship 2007; Pearl 2010) that coherently theorizes causal research. Just like

ordinary least squares analysts, QCA users may or may not violate that framework's principles. Thus, tracing the perhaps interesting relation between QCA and the counterfactual framework would provide no distinctive insight on QCA. Hence, despite the potential interest, space constraints preclude our engaging this issue.

6. Theologians (e.g., Kierkegaard [1844] 2009), philosophers (e.g., Byrne 1978), physicists (e.g., Einstein [1926] 2005), and writers (e.g., Tolstoy [1869] 1966) join sociologists in debating stochasticism (sometimes free will) versus determinism. We focus only on the former form.
7. A trivial explanation would have one independent variable for every case.
8. Equation (1) could be parametric even though QCA is a nonparametric method. Any method-data mismatch is irrelevant, because nonparametric methods can discern patterns in parametric data, whereas the opposite is not in general true and, more basically, because analysts always proceed while uncertain of whether their methods match the world fundamentally or not. QCA is used in that same environment; thus, it is fair to use parametric data to evaluate QCA.
9. We thank an anonymous reviewer for pointing out this implication.
10. The assumption may cost precision, as stochasticists estimate $K + \delta > K$ quantities.
11. Seawright (2005) considered the status of omitted variables in causal analysis in QCA. However, Ragin sometimes frames QCA as an exploratory technique (e.g., Ragin 2008a:90). Thus, we extend Seawright's result to other uses of QCA and to all deterministic methods.
12. $\Pr(Y_i = 1) = e^{(0.986 - 0.316 * X_1 + 2.486 * X_2 - 1.220 * X_3 - 1.786 * X_4) / (1 + e^{(0.986 - 0.316 * X_1 + 2.486 * X_2 - 1.220 * X_3 - 1.786 * X_4)})}$ is the equation; we then classified all cases such that $Y = 0$ if $\Pr(Y = 1) \leq .5$, and $Y = 1$ otherwise.
13. A linktest (Pregibon 1980) supported the symmetric link specification.
14. We presume that analysts use QCA to analyze data rather than only to summarize their claims about data. If so, analysts should explore the causal status of less expected causes, which we do here in this part of the analysis.
15. If we had treated AMP as causal for $Y = 1$, our qualitative conclusions would not change.
16. These expressions can be stated in terms of Rubin's causal model (Holland 1986). Such a treatment could explore whether study of unit-specific effects has implications for QCA; however, because of space limitations, we do not pursue that connection here.
17. The occupational scaling tradition has debated scaling a categorical variable; Ragin addresses how to scale continuous variables. Still, many of the theoretical issues are similar.
18. For example, debate on poverty measurement concerns how to base poverty thresholds in solid theory and feasible measurement (Citro and Michael 1995; Ragin 2008a:72). The use of a socioeconomic index (Duncan 1961), a measure of Marxist class position (Wright et al. 1982), or an index of job desirability (Jencks, Perman, and Rainwater 1988) turns partly on calibration (e.g., how many must one employ to be a capitalist?). The debate over earnings and log earnings clearly

- concerns calibration (e.g., Beck, Horan, and Tolbert 1978; Hauser 1980; Hodson 1985).
19. This was actually the first analysis we conducted. Knowing that our work could become a rote and therefore unrepresentative application of QCA, we had each author alone analyze one of two randomly allocated sets of cases. One author, faced with data as described, decided to use a .6 consistency threshold, a decision akin to retaining a variable with a p value of .056, perhaps because of theory. Unless QCA is meant to excise researcher judgment completely, such decisions seem to be the researcher's responsibility and may at times diverge from general advice. However, we reanalyzed both data sets using the alternative consistency threshold, to explore the implication of this difference as well as other matters. Thus, the ramification of the decision for this study is nonexistent, for if one views a threshold of .6 as too low, one may focus on the other results instead.
 20. Ganzeboom, Treiman, and Ultee (1991) show that after several illuminating studies that explicitly treated measurement error between 1965 and 1985, (e.g., Bielby, Hauser, and Featherman 1977; Hauser, Tsai, and Sewell 1983), stratification analysts moved to other modeling strategies and rarely account for measurement error. Hence, the measurement error criticism of QCA is apt, but a similar criticism applies widely.
 21. The coverage of these results, the lowest we obtained, remains substantially above that reported for some QCA studies deemed successful (e.g., Ragin 2008a:205). If coverage is like R^2 , then it informs but is not a criterion for accepting or rejecting results. Contra Schneider and Wagemann (2010:414), some analysts do not report coverage (e.g., Mahoney 2010), making it hard to assess whether our coverage is low, normal, or high.
 22. On the logic and history of probability sampling, see Kruskal and Mosteller (1980).

References

- Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6: 169–86.
- Abbott, Andrew and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96:144–85.
- Achen, Christopher H. 2005. "Two Cheers for Charles Ragin." *Studies in Comparative International Development* 40:27–32.
- Adams, Julia. 2005. *The Familial State: Ruling Families and Merchant Capitalism in Early Modern Europe*. Ithaca, NY: Cornell University Press.
- Amenta, Edwin and Drew Halfmann. 2000. "Wage Wars: Institutional Politics, WPA Wages, and the Struggle for U.S. Social Policy." *American Sociological Review* 65: 506–28.
- Beck, E.M., Patrick M. Horan, and Charles M. Tolbert II. 1978. "Stratification in a Dual Economy: A Sectoral Model of Earnings Determination." *American Sociological Review* 43:704–20.

- Berg-Schlosser, Dirk, Gisèle De Meur, Benoît Rihoux, and Charles C. Ragin. 2009. "Qualitative Comparative Analysis (QCA) as an Approach." Pp. 1–18 in *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Benoît Rihoux and Charles C. Ragin. Thousand Oaks, CA: Sage.
- Bielby, William T., Robert M. Hauser, and David L. Featherman. 1977. "Response Errors of Black and Nonblack Males in Models of the Intergenerational Transmission of Socioeconomic Status." *American Journal of Sociology* 82:1242–88.
- Burke, James. 1978. *Connections*. New York: Little, Brown.
- Byrne, Peter. 1978. "Miracles and the Philosophy of Science." *Heythrop Journal* 19: 162–70.
- Camic, Charles. 1986. "The Matter of Habit." *American Journal of Sociology* 91: 1039–87.
- Citro, Constance and Robert T. Michel. 1995. *Measuring Poverty: A New Approach*. Washington, DC: National Academies Press.
- Clark, William Roberts, Michael J. Gilligan, and Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14:311–31.
- Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74:829–36.
- Clogg, Clifford C. 1982. "Some Models for the Analysis of Association in Multiway Cross-classifications Having Ordered Categories." *Journal of the American Statistical Association* 77:803–15.
- Cronqvist, Lasse and Dirk Berg-Schlosser. 2006. "Determining the Conditions of HIV/AIDS Prevalence in Sub-Saharan Africa: Employing New Tools of Macro-qualitative Analysis." Pp. 145–66 in *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-qualitative Divide*, edited by Benoît Rihoux and Heike Grimm. New York: Springer.
- De Meur, Gisèle and Benoît Rihoux. 2002. *L'Analyse Quali-Quantitative Comparée (AQQC-QCA)*. Louvain-La-Neuve, Belgium: Academia Bruylant.
- Duncan, Otis Dudley. 1961. "A Socioeconomic Index for All Occupations." Pp. 109–38 in *Occupations and Social Status*, edited by Albert J. Reiss Jr. New York: Free Press.
- Duncan, Otis Dudley. 1979. "How Destination Depends on Origin in the Occupational Mobility Table." *American Journal of Sociology* 84:793–803.
- Duşa, Adrian. 2010. "A Mathematical Approach to the Boolean Minimization Problem." *Quality and Quantity* 44:99–113.
- Dy, Sydney M., Pushkal Garg, Dorothy Nyberg, Patricia B. Dawson, Peter J. Pronovost, Laura Morlock, Haya Rubin, and Albert W. Wu. 2005. "Critical Pathway Effectiveness: Assessing the Impact of Patient, Hospital Care, and Pathway Characteristics Using Qualitative Comparative Analysis." *Health Services Research* 40:499–516.
- Dyson, F. W., A. S. Eddington, and C. Davidson. 1920. "A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 220: 291–333.

- Einstein, Albert. (1926) 2005. "Letter to Max Born (4 December 1926)." P. 88 in *The Born-Einstein Letters: Friendship, Politics and Physics in Uncertain Times*, translated by Irene Born. New York: Macmillan.
- Feynman, Richard P. 1986. "Appendix F—Personal Observations on Reliability of Shuttle." Pp. F1–F5 in *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, Vol. 2, by William Rogers, Neil A. Armstrong, David C. Acheson, Eugene E. Covert, Richard P. Feynman, Donald J. Kutnya, Sally K. Ride, Robert W. Rummel, Joseph F. Sutter, Arthur B.C. Walker, Jr., Albert D. Wheelen, and Charles E. Yeager. Washington, DC: U.S. Government Printing Office.
- Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80:27–38.
- Fuchs, Stephan. 1997. "A Sociological Theory of Objectivity." *Science Studies* 11: 4–26.
- Fujita, Taisuke. 2009. "Developed and Democratic Countries' Policy-making on Dispute Settlement in the GATT/WTO: Exploring Conjunctural and Multiple Causations by Comparing QCA and Regression Analysis." *Sociological Theory and Methods* 24:181–202.
- Ganzeboom, Harry B. G., Donald J. Treiman, and Wout C. Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17:277–302.
- George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT Press.
- Gigerenzer, Gerd. 1991. "From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology." *Psychological Review* 98:254–67.
- Glaesser, Judith, Richard Gott, Ros Roberts, and Barry Cooper. 2009. "The Roles of Substantive and Procedural Understanding in Open-ended Science Investigations: Using Fuzzy Set Qualitative Comparative Analysis to Compare Two Different Tasks." *Research in Science Education* 39:595–624.
- Goldberger, Arthur S. 1991. *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
- Goldthorpe, John H. 2000. "Current Issues in Comparative Macrosociology." Pp. 45–64 in *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. New York: Oxford University Press.
- Gran, Brian, and Dawn Aliberti. 2003. "The Office of the Children's Ombudsperson: Children's Rights and Social-policy Innovation." *International Journal of the Sociology of Law* 31:89–106.
- Gribbin, John. 1984. *In Search of Schrödinger's Cat: Quantum Physics and Reality*. Toronto, Canada: Bantam.
- Grofman, Bernard, and Carsten Q. Schneider. 2009. "An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression." *Political Research Quarterly* 62:662–72.
- Hauser, Robert M. 1980. "Comment on Beck, Horan, and Tolbert, 'On Stratification in a Dual Economy.'" *American Sociological Review* 45:702–12.
- Hauser, Robert M. and Douglas K. Anderson. 1991. "Post-High School Plans and Aspirations of Black and White High School Seniors: 1976–86." *Sociology of Education* 64:263–77.

- Hauser, Robert M., Shu-Ling Tsai, and William H. Sewell. 1983. "A Model of Stratification with Response Error in Social and Psychological Variables." *Sociology of Education* 56:20–46.
- Hauser, Robert M. and John Robert Warren. 1997. "Socioeconomic Indexes for Occupations: A Review, Update, and Critique." *Sociological Methodology* 27: 177–298.
- Heinze, Georg and Michael Schemper. 2002. "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409–19.
- Herrnstein, Richard J. and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hodson, Randy. 1985. "Some Considerations Concerning the Functional Form of Earnings." *Social Science Research* 14:374–94.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945–60.
- Hollingsworth, Rogers, Robert Hanneman, Jerald Hage, and Charles Ragin. 1996. "The Effect of Human Capital and State Intervention on the Performance of Medical Systems." *Social Forces* 75:459–84.
- Hosmer, David W. and Stanley Lemeshow. 1989. *Applied Logistic Regression*. New York: John Wiley.
- Hyötyniemi, Heikki. 2006. "Neocybernetics in Biological Systems." Report 151. Helsinki, Finland: Control Engineering Laboratory, Helsinki University of Technology.
- Jencks, Christopher, Lauri Perman, and Lee Rainwater. 1988. "What Is a Good Job? A New Measure of Labor Market Success." *American Journal of Sociology* 93: 1322–57.
- Kant, Immanuel. (1781) 2008. *The Critique of Pure Reason*. Translated by Max Müller and Marcus Weigelt. New York: Penguin Classics.
- Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton*. Boston: Houghton Mifflin.
- Kierkegaard, Søren. (1844) 2009. *Philosophical Fragments*. New York: Feather Trail.
- Kruskal, William and Frederick Mosteller. 1980. "Representative Sampling, IV: The History of the Concept in Statistics, 1895–1939." *International Statistical Review* 48:169–95.
- Laumann, Edward O., Peter V. Marsden, and David Prensky. 1983. "The Boundary Specification Problem in Network Analysis." Pp. 18–34 in *Applied Network Analysis: A Methodological Introduction*, edited by Ronald S. Burt, et al. Beverly Hills, CA: Sage.
- Lazersfeld, Paul. 1949. "The American Soldier—An Expository Review." *Public Opinion Quarterly* 13:377–404.
- Liao, Tim Futing. 2001. "Fuzzy-set Social Science." *Social Forces* 80:354–56.
- Liebertson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.
- Liebertson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70:307–20.

- Lieberson, Stanley. 1994. "More on the Uneasy Case for Using Mill-type Methods in Small-*N* Comparative Studies." *Social Forces* 72:1225–37.
- Lieberson, Stanley. 2001. "A Review of *Fuzzy-set Social Science* by Charles C. Ragin." *Contemporary Sociology* 30:331–34.
- Lieberson, Stanley. 2004. "Comments on the Use and Utility of QCA." *Qualitative Methods* 2:13–14.
- Longest, Kyle C. and Stephen Vaisey. 2008. "fuzzy: A program for performing Qualitative Comparative Analyses (QCA) in Stata." *Stata Journal* 8:79–104.
- Lucas, Samuel R. 2012. "The Road to Hell . . . : The Statistics Proposal as the Final Solution to the Sovereign's Human Rights Question." *Wisconsin International Law Journal* 30:259–343.
- Lucas, Samuel R. 2014a. "Beyond the Existence Proof: Ontological Conditions, Epistemological Implications, and In-depth Interview Research." *Quality and Quantity* 48:387–408.
- Lucas, Samuel R. 2014b. "An Inconvenient Dataset: Bias and Inappropriate Inference with the Multilevel Model." *Quality and Quantity* 48:1619–49.
- Mahoney, James. 1999. "Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis." *American Journal of Sociology* 104:1154–96.
- Mahoney, James. 2000. "Strategies of Causal Inference in Small-*N* Analysis." *Sociological Methods and Research* 28:387–424.
- Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41:412–36.
- Mahoney, James. 2010. *Colonialism and Postcolonial Development: Spanish America in Comparative Perspective*. New York: Cambridge University Press.
- Mahoney, James and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14:227–49.
- Mann, Michael. 2005. *The Dark Side of Democracy: Explaining Ethnic Cleansing*. New York: Cambridge University Press.
- Manski, Charles F. 1995. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press.
- Marx, Axel and Geert van Hootegeem. 2007. "Comparative Configurational Case Analysis of Ergonomic Injuries." *Journal of Business Research* 60:522–30.
- Mill, John Stuart. (1843) 1906. *A System of Logic: Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. New York: Longmans, Green.
- Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- Neter, John, William Wasserman, and Michael H. Kutner. 1989. *Applied Linear Regression Models*, 2nd ed. Homewood, IL: Irwin.
- Pearl, Judea. 2010. "The Foundations of Causal Inference." *Sociological Methodology* 40:75–149.
- Perrow, Charles. 1999. *Normal Accidents: Living with High-risk Technologies*. Princeton, NJ: Princeton University Press.
- Pregibon, Daryl. 1980. "Goodness of Link Tests for Generalized Linear Models." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29:15–24.

- Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, Charles C. 1992. "'Casing' and the Process of Social Inquiry." Pp. 217–26 in *What Is a Case? Exploring the Foundations of Social Inquiry*, edited by Charles C. Ragin and Howard S. Becker. New York: Cambridge University Press.
- Ragin, Charles C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.
- Ragin, Charles C. 2005. "Core versus Tangential Assumptions in Comparative Research." *Studies in Comparative International Development* 40:33–38.
- Ragin, Charles C. 2006. "The Limitations of Net-effects Thinking." Pp. 13–41 in *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide*, edited by Benoît Rihoux and Heike Grimm. New York: Springer.
- Ragin, Charles C. 2008a. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.
- Ragin, Charles C., with Sarah Ilene Strand and Claude Rubinson. 2008b. *User's Guide to Fuzzy Set / Qualitative Comparative Analysis*. Tucson, AZ: Charles Ragin and Sean Davey.
- Ragin, Charles C. 2009. "Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA)." Pp. 87–122 in *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Benoît Rihoux and Charles C. Ragin. Thousand Oaks, CA: Sage.
- Ragin, Charles C. and Benoît Rihoux. 2004. "Qualitative Comparative Analysis (QCA): State of the Art and Prospects." *Qualitative Methods* 2:3–13.
- Reichenbach, Hans. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.
- Rihoux, Benoît and Heike Grimm. 2006. *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-qualitative Divide*. New York: Springer.
- Rihoux, Benoît and Charles C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks, CA: Sage.
- Robison, Wade, Roger Boisjoly, David Hoeker, and Stefan Young. 2002. "Representation and Misrepresentation: Tufte and the Morton Thiokol Engineers on the Challenger." *Science and Ethics* 8:59–81.
- Rogers, William, Neil A. Armstrong, David C. Acheson, Eugene E. Covert, Richard P. Feynman, Donald J. Kutnya, Sally K. Ride, Robert W. Rummel, Joseph F. Sutter, Arthur B.C. Walker Jr., Albert D. Wheelen, and Charles E. Yeager. 1986. *Report of the Presidential Commission on the Space Shuttle Challenger Accident*. Washington, DC: U.S. Government Printing Office.
- Savolainen, Jukka. 1994. "The Rationality of Drawing Big Conclusions Based on Small Samples: In Defense of Mill's Methods." *Social Forces* 72:1217–24.
- Schneider, Carsten Q. and Claudius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-sets." *Comparative Sociology* 9:397–418.
- Seawright, Jason. 2005. "Qualitative Comparative Analysis vis-à-vis Regression." *Studies in Comparative International Development* 40:3–26.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.

- Somers, Margaret. 2008. *Genealogies of Citizenship: Markets, Statelessness, and the Right to Have Rights*. New York: Cambridge University Press.
- Steinmetz, George. 2007. *The Devil's Handwriting: Precoloniality and the German Colonial State in Qingdao, Samoa, and Southwest Africa*. Chicago: University of Chicago Press.
- Stouffer, Samuel A. 1949. *Studies in Social Psychology in World War II: The American Soldier*. Princeton, NJ: Princeton University Press.
- Thompson, Linda and Alexis J. Walker. 1982. "The Dyad as the Unit of Analysis: Conceptual and Methodological Issues." *Journal of Marriage and the Family* 44: 889–900.
- Tilly, Charles. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Russell Sage.
- Tolstoy, Leo. (1869) 1966. *War and Peace: The Maude Translation, Backgrounds and Sources, Essays in Criticism*. New York: W.W. Norton.
- Tufte, Edward. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*. Cheshire, CT: Graphics Press.
- Vaisey, Stephen. 2007. "Structure, Culture, and Community: The Search for Belonging in 50 Urban Communes." *American Sociological Review* 72:851–73.
- Vaughan, Diane. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University of Chicago Press.
- Wagemann, Claudius and Carsten Schneider. 2010. "Qualitative Comparative Analysis (QCA) and Fuzzy-sets: Agenda for a Research Approach and a Data Analysis Technique." *Comparative Sociology* 9: 376–96.
- Wright, Erik Olin, Cynthia Costello, David Hachen, and Joey Sprague. 1982. "The American Class Structure." *American Sociological Review* 47:709–26.
- Wu, Lawrence. 2003. "Event History Models for Life Course Analysis." Pp. 477–502 in *Handbook of the Life Course*, edited by Jeylan T. Mortimer and Michael J. Shanahan. New York: Elsevier.
- Xie, Yu. 1992. "The Log-multiplicative Layer Effect Model for Comparing Mobility Tables." *American Sociological Review* 57:380–95.

Author Biographies

Samuel R. Lucas is a faculty affiliate of the Institute for Research on Poverty of the University of Wisconsin–Madison and a professor of sociology at the University of California, Berkeley, with research interests in four areas: social stratification, the sociology of education, research methods, and statistics. His first book, *Tracking Inequality: Stratification and Mobility in American High Schools*, received the 2000 Willard Waller Award. *Theorizing Discrimination in an Era of Contested Prejudice* (2008), which develops a coherent definition of discrimination, and *Just Who Loses?* (2013), which uses that definition to construct indicators of discrimination and estimates effects of race and sex discrimination on 13 outcomes over the life course for targets and nontargets of discrimination in the United States, provide the first two volumes of his three-volume analysis of discrimination. In addition to work on stratification theory, his most recent methods papers include a *Wisconsin International Law Journal* article critically evaluating the use of statistics in human rights adjudication and two papers,

one on in-depth interviewing and one on multilevel modeling, both in *Quality and Quantity*.

Alisa Szatrowski is a doctoral candidate in the Department of Sociology at the University of California, Berkeley. She holds a master's degree in sociology from the University of California, Berkeley (2010), and was a recipient of the chancellor's fellowship. Her work focuses on stratification, poverty, research methods, education, and health using both qualitative and quantitative methodologies. Her master's thesis examined the impact of poverty on later-life health focusing on the relative importance of timing, duration, and change in income status to health outcomes. She is currently working on her dissertation research, a comparative ethnographic project on high schools that are academically successful with low-income students. She has also worked as a consultant to design data-gathering materials for research on teacher effectiveness.