REJOINDER: TAKING HEAT AND GIVING LIGHT—REFLECTIONS ON THE EARLY RECEPTION OF "QUALITATIVE COMPARATIVE ANALYSIS IN CRITICAL PERSPECTIVE"

*Samuel R. Lucas**

*University of California, Berkeley
Corresponding Author: Samuel R. Lucas, Lucas@berkeley.edu

1. INTRODUCTION

Many years ago, my sister told me of a disagreement she had with an acquaintance. She decided to read over the phone to me emails she and her acquaintance had exchanged. After she had read a few I conveyed to her that every time she read her email she used a sincere and gentle voice, but every time she read her acquaintance's email she used a rough and angry voice. We laughed about that, she reread the messages to me in a neutral voice, and although she and her acquaintance still disagreed, the rereading created room for an amicable resolution.

Juxtaposing the comments from Bowers, Collier, Olsen, and Seawright, on the one hand, with those from Fiss, Marx, and Rihoux (FMR); Ragin; and Vaisey, on the other, it sometimes seems as if authors in the latter set read our paper with a rough and angry voice pounding in their heads.

I ask the reader to use a neutral voice when reading our paper and in reading this rejoinder. Whatever affect I experienced while preparing this rejoinder, a sad voice, not an angry one, is most consistent with it.

The symposium neglected some potentially useful results; thus, I first briefly note a few such findings. Next, I try to reorient the symposium by addressing the latter set of authors' claims about tone and the quantitative perspective they impute to us. I then discuss Bowers's, Collier's, and Seawright's comments, turning next to general claims of qualitative

comparative analysis (QCA) proponents. Few comments concern Olsen's work because her work is farther afield, and space constraints preclude serious treatment of it. However, she raises intriguing issues of the grounds for social analysis. Some of her views are not that far from my own. Yet, I see no value to attaching her grounding to QCA.

Next, the replication efforts Ragin, FMR, and Vaisey offer are assessed. These analysts were unable to replicate some of our work, suggesting serious errors on our part. Thus, three observations must be made at the outset: (1) Ragin's analyses have key errors that explain his failures; (2) FMR provide insufficient documentation to assess whether their analyses are true replications, but they dismiss most of our solutions on contradictory reasoning; and (3) Vaisey's Stata code shows that he changed key aspects of our analyses, derailing much of his replication effort. Although all interpret their work as salvaging QCA, with a mixture of errors, incomplete documentation, and changes in method, that they replicate some but not all of our findings should not surprise. As will become clear below, their replication failures do not invalidate our findings.

I conclude with summary observations on scholarly discourse. But I begin by noting selected important findings neglected in the symposium.

2. SELECTED USEFUL CONCLUSIONS POSSIBLY LOST IN THE DEBATE

The proof that necessary and sufficient causes are versions of functional form asymmetry (Lucas and Szatrowski, this volume, pp. 35–36) is a potentially interesting result that further erodes the quantitative/qualitative divide.

Placing QCA's method of finding causal asymmetry in the context of a discussion on identification (this volume, pp. 27–38) is potentially useful, for it draws attention to the need of macrocomparative analyses to address identification issues. This development also further erodes the qualitative/quantitative divide. Showing that QCA's procedures exacerbate problems posed by measurement error is also a contribution. Noting the arbitrariness of the preference for the logistic distribution in calibrating variables for fuzzy set analysis is potentially worthwhile, but it is of perhaps less scope than the aforementioned elements.

The paper may contain other potentially useful contributions. Had the symposium assessed these, perhaps some would have survived, while others would not have. Thus, I can only hope that any useful nuggets will be found at some point and that any errors will not impede our study of the social world.

3. ON THE CONDUCT OF INTELLECTUAL DISAGREEMENT

3.1. *Tone*

The following were drawn from three of the seven commentaries (all italics were added for highlighting):

1. "At first glance, it *seemed* as if L&S wanted to be part of building these bridges" (Vaisey, this volume, p. 111).
2. "In addition, we regret their aggressive and negative writing style" (FMR, this volume, p. 98).
3. "The *stated* goal of Lucas and Szatrowski [this volume; hereafter L&S] is to provide a critical evaluation of qualitative comparative analysis (QCA)" (Ragin, this volume, p. 80).
4. "To calibrate the two pressure variables as fuzzy sets, they *appear* to use 50 = .05, 100 = .27, and 200 = .95" (Ragin, this volume, p. 89).
5. "To calibrate temperature as a fuzzy set, they *appear* to use 32°F = .05, 65°F = .50, and 98°F = .95" (Ragin, this volume, pp. 89–90).
6. "Finally, it is important to *have full disclosure* and *avoid giving a one-sided* presentation" (Ragin, this volume, p. 91).
7. "What I am referring to is the simple fact that L&S *failed to present* the logistic regression analysis of this data set, which is surprising given that they often use this technique in their critique" (Ragin, this volume, pp. 91–92).
8. "They botch their only application of QCA to empirical evidence . . ." (Ragin, this volume, p. 93).
9. "They . . . *fail to disclose* the results of a logistic regression analysis of the same data" (Ragin, this volume, p. 93).

These excerpts come from commentaries that criticized our paper's alleged tone. Given their response, I am not optimistic a productive dialogue can flow from the symposium. But, perhaps real dialogue can be

salvaged. To begin a salvage effort, it may help to directly convey my motives for studying QCA.

I learned of crisp-set QCA sometime between 1999 and 2001, when a sociology student proposed using the method in their MA paper. I was intrigued. I thought that this method could, if developed in a certain direction, resolve a problem in my research area. My research on high school tracking used nationally representative data on approximately 1,000 schools, each school with about 15 probability-sampled students. I wondered whether it would be possible to simultaneously conduct a QCA in each school context and use the school-specific causal recipes as dependent variables in a school-level analysis. For example, perhaps different schools had different causal recipes for students' entrance to demanding classes, and perhaps school factors (e.g., racial diversity, class diversity, material resources) might be correlated with those recipes. Substantively, schools might fall into one of a small set of regimes—sets of causal recipes—and their regime may be partly determined by school-level factors.[1] If so, standard multilevel modeling might never discover that aspect of reality. To excavate such patterns, I hoped to develop multilevel QCA (mlQCA).

Alas, other assignments and responsibilities prevented me from pursuing this work for several years, but in spring 2007, my calendar was cleared. To construct mlQCA I needed to deepen my understanding of QCA, so I began reading. I quickly realized that in the six to eight years since my introduction to QCA, important changes (e.g., the development of fuzzy-set QCA [fsQCA]) had occurred and major controversy had erupted. While reading, I eventually encountered Lieberson's (2004) proposal for simulation studies, which resonated with me because at the time, I was working on a simulation study on sampling and multilevel models (Lucas 2014) to assess some issues relevant to the second volume in my discrimination series (Lucas 2013). So, I recruited a student into the project, and we set out.

Let me be clear: I would *never* have recruited a student into the project had I expected to find no value to the method. Based on Ragin's substantive published work, claims about QCA, and reputation as a top sociologist, as well as the gathering of other notable scholars around the method, I expected to find some ways the method worked and some limitations. After publishing a simulation study, I expected to turn to devising mlQCA. I was excited by the prospect of eventually contributing a new method of analysis while answering key questions in a substantive

domain of interest to me. Regrettably, it did not work that way—as we read the QCA literature more closely, what had been a promising line became less and less promising. Once we began to obtain our analysis results, I became persuaded that the method is seriously flawed. I therefore dropped the idea of devising mlQCA.

One complaint is that we mischaracterize QCA. Space constraints prevented us from outlining every opinion of each QCA proponent, but we tried to describe QCA accurately. We had no desire to caricature. Indeed, how could we caricature QCA? The paper underwent five rounds of external review involving at least one pro-QCA reviewer. This stringent review process likely removed any errors. And, if there are errors, they are honest ones based on our reading of QCA proponents; if handled well, their articulation could help QCA proponents to clarify their view. However, disagreement with QCA proponents does not imply we misunderstand their claims.

As for our paper's tone, I genuinely do not believe it is aggressive, but I am willing to hear evidence on that point because I take seriously any allegation about my work as I seek to improve my understanding and contribute to the public understanding of society *and* sociology. However, I admit our paper is direct. For example, we found several places where the method did not behave as claimed. These findings led us to subtitle our concluding section "Epistemological Bait and Methodological Switch." This is an admittedly direct, and some might say provocative, subtitle. It is, at the least, possibly a memorable subtitle. Even so, we did not question the intent or integrity of QCA proponents; we simply clearly conveyed our position—in our estimation, the method does not do what it purports to do. A productive response either would show that there is no disjuncture between claims and actual operation of QCA or would repair any disjunctures. Protesting the paper's alleged tone does not add one character of code to QCA algorithms that we and an increasing number of others have shown do not work.

Indeed, despite claims that the paper has an aggressive and negative tone, we never suggested anyone was "failing to disclose"; never suggested any person possessed a "stated goal" (which might imply we believe they have some other *hidden* goal); never contended anyone offered a "one-sided presentation"; never claimed someone "appeared to" do something that they explicitly, clearly, write that they did—because using the language of "appears" implies one is not really sure the author did what they said they did. Yet, our paper has inspired

exactly these kinds of responses at the same time as the authors of those responses claim that our paper is "aggressive and negative." The response of those QCA proponents is disappointing. Years ago, Mahoney (2003:21), a QCA user, took the rare step of reminding QCA proponents in print that critics "may raise real challenges to fs/QCA, and advocates of fs/QCA need to avoid responding to them with a knee-jerk defensiveness" (Mahoney 2003:21), suggesting that this is not the first time that critique of QCA has elicited the kind of response our paper has received.

Sadder still, we held Ragin—main architect of the method—in high regard, and we expressed this position in early submitted versions of the paper. For example, the abstract to our original paper (Lucas and Szatrowski 2011) included the following:

> Ragin (1987, 2000, 2008) and colleagues have tirelessly worked what appeared to be a promising line of methodological possibility, bringing it to relative maturity for broader use and evaluation. Only if scholars are willing to take such risks may we collectively push the boundaries of inquiry. Thus, their contribution is laudable. Still, our evaluation of the method has a clear and unfortunate conclusion—analysts should refrain from using QCA/ fsQCA in favor of existing effective methods. (p. ii)

In the introduction to that same manuscript, we further wrote,

> QCA/fsQCA failed to reproduce known answers for several datasets, an exercise we began fully expecting multiple sub-studies would map the value and limitations of the method. However, in conducting these analyses we came to realize that what the method is *understood* to do, and what its inner workings *actually* do, are very different. We believe this disjuncture is inadvertent and widely unrecognized, not owing to deceit on the part of QCA/ fsQCA architects. (p. 2; italics in original)

And, we closed that manuscript as follows:

> We sincerely commend Ragin and colleagues for their ambitious efforts. Only with such efforts can the discipline fully explore the analytic possibilities and the limits thereof. Alas, when we evaluate the method we find it and its logic cannot be sustained. Thus, we conclude that, despite the hope it represented, researchers should abandon QCA/fsQCA. (p. 85)

Regrettably, a pro-QCA reviewer claimed that our manuscript was sarcastic and that our *real* motive was to attack QCA. Such claims, plus

the ever-present issue of length, combined during review to force us to remove the text above.

Thus, a project begun owing to a desire to extend a method, a project that could not exist did we not hold in high regard the creators of the method on which our work would stand, culminates in our words being discounted (e.g., "they appear to use"; Ragin, this volume, p. 89), while our motives (e.g., "it seemed as if L&S wanted"; Vaisey, this volume, p. 111) and integrity (e.g., "they fail to disclose"; Ragin, this volume, p. 93) are challenged. This is very disappointing.

3.2. *Attributions: Do I Stand Where QCA Proponents Claim I Stand?*

Ragin (this volume) correctly claims that evaluations must be sensitive to the aims of the method being evaluated. The question to ask is whether our standards are fair. Unfortunately, some seem to determine the validity of evidence in line with foundational commitments they impute to those who produce the evidence. Sadly, because QCA proponents' imputations can motivate some to dismiss our findings, the imputations require response.

I and my coauthor have been said to champion "conventional applications of quantitative methods" (Ragin, this volume, p. 82) and to come from "a 'quant' perspective" (Vaisey, this volume, p. 108). Is this correct?

Consider the actual paper. In discussing asymmetric causation and identification, we note that "Lieberson's (1985) [English-language dominance] example introduced information on event sequencing" (Lucas and Szatrowski, this volume, p. 37), and we encourage narrative analysis, process tracing, and sequence analysis as methods for assessing dynamic asymmetric causation. Later, in suggesting general methods for those with questions they might have tried to answer using QCA, we explicitly offer narrative strategies and approvingly cite several nonquantitative comparative historical analyses. How can one maintain that we champion conventional quantitative analysis when we explicitly embrace narrative analysis, an approach arguably opposite conventional quantitative analysis?

Second, Vaisey (this volume, p. 108) attributes to us a "quantitative perspective" from which our analysis is supposed to derive. I ask, what is the quantitative perspective we are supposed to have adopted? Do we

have a set of cases (as opposed to one case)? Yes, and so does every QCA user. Are our samples probability samples? Sometimes yes, and sometimes no, as is the case for QCA users. Did we seek to generalize from studied cases to the population of cases? Usually no, as is the case for many QCA users. In short, what in the paper indicates it is based on a "quantitative perspective"? True, much of our research is largely quantitative (e.g., Lucas 1999, 2013). Yet, our work is broader than Vaisey's imputation suggests; we have historical/theoretical (e.g., Lucas 2008), formal theorizing (Lucas 2009), and ethnographic (Szatrowski forthcoming) work. Further, our work has critiqued statistical practice, from showing that some popular statistical model/data combinations are indefensible (Lucas 2014) to directly opposing use of statistical methods at all for some questions, championing conventional *qualitative* research methods instead (Lucas 2012). Some of these works may be less well known than the quantitative work, but they exist and join the content of the paper in contradicting QCA proponents' attributions. Intriguingly for those who impute commitments, our paper can be understood as championing conventional, nonformal, *qualitative* research over the formalization QCA represents.

Some readers might doubt the ability of champions of conventional quantitative analysis to assess QCA. Whatever merits such reasoning may hold, the reasoning does not apply here, for we are not those champions.

Of course, all authors have a standpoint. The foundation of my standpoint is a small set of claims that are ultimately grounded in the belief that there are social-world constraints with *transmethodological* implications. By attending to those constraints, analysts may forge data collection and data analytic methods that at least partially loosen the bonds of those constraints and allow us to learn something of how the social world operates. To do so, however, we must make some assumptions, so objectivity is impossible. Yet, systematicity is possible, is extremely useful for furthering dialogue, and comes in varied forms depending on the research technique. Still, systematicity can be pursued too far— some phenomena just cannot be studied effectively with formal methods. For studying the social world, sound methods that use numbers and sound methods that do not use numbers have been developed. More may be developed. Even so, a proposed method must pass key, demanding tests, including tests for coherence, before a discipline can justifiably accept it. Such tests are necessary because analytic methods are

technologies we use to reveal aspects of the world to us. We will constantly misunderstand our world if we ease or eliminate our testing of methods because, as Feynman (1986:F5) indicates, "for a successful technology, reality must take precedence over public relations, for nature cannot be fooled." I accept Feynman's maxim.

I do not see how these commitments invalidate our analyses.

3.3. *Concluding Remarks on the Conduct of Intellectual Disagreement*

The response of most QCA proponents to our paper is disappointing not because of the treatment of our paper but, instead, because this is not only about the treatment of our paper. It is, more fundamentally, about the treatment of disagreements among scholars. Scholarship is seriously endangered when critiques elicit responses such as those provided by three of the four pro-QCA commentaries. Some may see no difference between the directness of our paper and QCA proponents' responses. However, there is a world of difference between directly articulating differences of opinion, perhaps in a colorful or memorable way, and subtly or not-so-subtly questioning the motives of others, implying deception by others, and/or imputing fundamental commitments to others as if those imputations invalidate others' research. Especially troubling is an allegation of deception for, implied or stated directly, such allegations are much more serious than being accused of being in error on some methodological point. Any intimation of deception implies one is unethical; even direct allegations of error imply only that one is human. Scholarly debate should be imbued with the latter understanding—everyone makes mistakes—and should banish the former in the absence of specific, documented, high-quality, nonpartisan evidence in support of the claim. Otherwise, the risk of being accused of deception—not to mention having one's motives questioned—can chill scholarly dialogue. All lose if that occurs.

Given the response of most QCA proponents to our paper, at this point I remain discouraged, and I observe: This is *no way* to engage in scholarship.

In the next section, I address the comments of Bowers, Collier, and Seawright.

4. DISPASSIONATE ASSESSMENTS OF QCA IN CRITICAL PERSPECTIVE

Bowers, Collier, and Seawright provide important contributions to the unfolding dialogue on QCA. The analyses of Bowers and Seawright indicate QCA is untrustworthy, and Collier takes the untrustworthiness of QCA as the point of departure for suggesting what to do in the wake of the emerging evidence.

Collier's response is consistent with belief in a unity of method. His call for simpler analyses resonates with work on graphical causal models (e.g., Pearl 2010; Elwert 2013). His call for greater attention to the truth table and avoiding analyzing it with QCA echoes Lieberson (2004). I agree with each of these calls, although it should be noted that while QCA may popularize truth tables in our fields, truth tables predate QCA (von Wright 1955).

Bowers (this volume) conducts two "method games," one easy, one difficult. He reports that QCA finds the true causal story 18 percent and 0 percent of the time, respectively. These findings are consistent with those we obtained. In response, Bowers suggests additional work on methods based on or derived from machine learning. I am not familiar enough with those methods to offer a considered response, but the general direction is the opposite of what I would suggest. Elsewhere I have argued that standardization is not always desirable; in that case, I showed problems with the shift from qualitative to quantitative analyses in human rights adjudication (Lucas 2012). Indeed, decades of research (e.g., Whyte 1943; Anderson 1976; Lareau 1989; Adams 2005) demonstrate the tremendous value of systematic, yet unstandardized, qualitative research. Although I am open to being persuaded of the value of machine learning and other means of standardizing qualitative methods for basic research, even if we are all persuaded, we must also recognize the value of time tested, less standardized (and possibly unstandardizable) techniques from which great knowledge has been produced. There is room for both, or, at least, there should be, as long as all analysts maintain a critical posture even toward their preferred methods.

In another simulation, Seawright (this volume) focuses on situations where the space carved out by the variables under study has zones without data. He constructs his data to match what he regards as QCA ontological assumptions, which he terms deterministic data, Boolean functional form that explains all cases, and no measurement error.

Under these conditions Seawright finds that only 14 percent of the time does QCA correctly identify the causes *and* exclude the noncausal factor. His findings are consistent with ours.

However, Seawright claims that in considering our work, "QCA supporters may complain with some justification that the simulations simply do not take their ontological commitments seriously" (Seawright, this volume, p. 119). Our analyses used both deterministic and stochastic data and varied several other features across the many analyses (functional form, overdetermined data, probability versus nonprobability selection of cases, and more) in an attempt to assess QCA under conditions acceptable to QCA proponents. Despite the diverse ontologies used, judging by the responses in the symposium, not one of those efforts was acceptable to QCA proponents.

Some QCA proponents seriously erred in their replications (see below). Other QCA proponents maintain one cannot evaluate QCA with simulated data. Complaints on those bases do not seem to be justified.

5. GENERAL ISSUES

QCA has been contentious partly because QCA proponents often distinguish QCA by relabeling concepts that many methods also employ, assert claims about QCA that no one could ever sustain, and reject key evaluative tools for assessing QCA, violating Feynman's maxim. The first makes QCA seem unique; the second makes QCA seem heroic; the third makes QCA seem unassailable. If an author rebuts their claims, some QCA proponents seem to dig in their heels and assert that the author misunderstands QCA. If confusion about QCA is to be eliminated, such conflict must be resolved. I will address the following issues the comments raised: (1) set-theoretic research, (2) configurations, (3) asymmetric causality, (4) case-specific explanations, and (5) simulated data.

5.1. *Set-theoretic Research: A Relabeled Concept*

Ragin (this volume) repeats again his oft-made claim that QCA is set-theoretic. In an earlier draft, we took up this issue, but we deleted that analysis owing to space. Ragin's raising the issue here provides an opportunity to consider the matter.

One argument is that social theories are often stated in set-theoretic terms. Yet, social analysts work with the tools available during their

career, and many later theorists follow earlier examples. As analysts theorized long before maturation of probabilistic frameworks and multivariate techniques, many theories may not reflect use of those tools. Still, set-theoretic terms may be tough to interpret or pose other problems for some work. Thus, I will show that avoiding contemporary tools (e.g., probabilities) is an unnecessary sacrifice to ask of empirical analysts.

Ragin (2008:88; italics in original) claims the set-theoretic conception fundamentally differs from others, contending that

> set membership scores that result from these transformations (ranging from 0.0 to 1.0) are *not* probabilities, but instead should be seen simply as transformations of interval scales into degree of membership in the target set. In essence, a fuzzy membership score attaches a *truth value*, not a probability, to a statement. . . . The difference between a truth value and a probability is easy to grasp, and it is surprising that so [sic] many scholars confuse the two. For example, the *truth value* of the statement "beer is a deadly poison" is perhaps about 0.05—that is, this statement is almost but not completely out of the set of true statements, and beer is consumed freely, without concern, by millions and millions of people every day. However, these same millions would be quite unlikely to consume a liquid that has a 0.05 *probability* of being a deadly poison, with death the outcome, on average, in one in twenty beers.

First, distinguish actual truth and perceived truth. Analysts have no access to the former except, perhaps, partly through the latter. Indeed, some doubt the existence of truth *qua* truth (e.g., Nietzsche [1873] 1976). These complexities mean that Ragin's purported ability to peg truth value separate from probability must be established, not simply asserted.

A prerequisite to establishing this ability is demonstrating probability/truth value distinctiveness, and this requires one to select a definition of probability. Ragin (2008) translates the .05 probability into a frequency ratio—1 in 20 beers is deadly—thus revealing his reliance on the "long-run frequencies" interpretation of probability. Some statistical researchers reject the long-run frequencies interpretation because it has its own problems, implying, for example, that probabilities of nonrepeatable events (e.g., the extinction of humans by the year 2100 CE) are inestimable. Some definitions of probability resolve these and other problems with the frequency interpretation and render fuzzy logic unnecessary (Cheeseman 1985). Indeed, many Bayesians define probability

as degrees of belief. Distinguishing the perceived truth value of, and the perceived degree of belief in, a statement is tough. Absent a distinction, truth values seem to offer nothing that Bayesian probabilities do not provide.

Ragin means the deadly beer example to distinguish truth values and probabilities. Although memorable, the illustration does not establish

$$\text{Probability}(\theta) \neq f(\text{TruthValue}(\varphi)), \qquad (1)$$

where f is a function. Ragin (2008) essentially maintains that truth value and probability are not related 1:1, but 1:1 is only one kind of relation. For example, both R^2 and adjusted R^2 are on the unit interval, yet, while their values are generally unequal, mathematical formulae translate from one to the other. Thus, even if one proved that probability .05 $\neq$ .05 truth value, that would not mean that probability and truth value are fundamentally distinct; consequently, Ragin's (2008) poisonous beer example fails to prove the veracity of expression (1).

If the concepts are dependent, then their dependence renders the use of either a *strategic* matter for a given study, not a matter of principle. As long as matters are appropriately translated into either terms, all is well.

Thus, returning to the example of the possibly poisonous pilsner, if .05 probability really translates to truth value .05, then I am confident that if a numerate consumer were told that there is a .05 truth value to the statement that beer is a deadly poison, the consumer would not purchase the beer. If a numerate consumer would purchase the beer, then that means truth value and probability are related in some way other than 1:1. If we determine that relation and equate truth value to a .05 probability, numerate consumers should reject the beer. Accordingly, I see little necessity for laborious articulation of factors in the language of sets, though the terminology is not harmful.

5.2. *Configurations: A Relabeled Concept*

The concept of interaction is simple—the effect of one factor (e.g., X_1) depends on the value of another (e.g., X_2). Interactions are typically multiplicative in the ordinary least squares (OLS) model, but ANOVA clarifies that the concept of interactions is not multiplicative. Further, nonmultiplicative interactions are possible in the OLS model.

Consider mothers' and fathers' years of education, each ranging from 0 to 22. One could multiply the two variables to specify an interaction

and place all three variables in a model. Alternatively, one could construct 529 binary variables (0 years of schooling for Mom, 0 years of schooling for Dad; 0 years of schooling for Mom, 1 year of schooling for Dad; and so on) and use all but one in a model. Both specifications allow the mother's education effect to differ depending on the father's education level, and thus both specifications estimate interaction effects. The binary variable specification is flexible enough to allow for threshold effects.

In the above example, I contrast completely multiplicative and completely disaggregated specifications. One may specify an interaction model between these extremes. For example, one might collapse all combinations where one parent has 8 years of schooling or less into one category. Doing so hypothesizes homogenous effects for children with a parent with less than 9 years of schooling regardless of the other parent's education. All other effects of mother's education differ owing to father's education and vice versa. Many such complex specifications are possible.

QCA proponents confuse one parametric specification of the concept of interaction with the concept itself, leaving no label for the general case of X_1 effects depending on the value of X_2. Then, another term—*configurations*—is adopted to cover the conceptual terrain now scoured of its usual label. However, as shown above, configurations and interactions are the same.

5.3. *Asymmetric Causality: An Unsustainable Claim*

Estimate a multinomial logit model for a trichotomous variable. Now, change the baseline category of the dependent variable to reestimate the model. Coefficient values will differ; predicted probabilities will not. Vaisey basically states that if you change an identifying assumption, you obtain nominally different results, which is true. However, you have no new knowledge; it only *seems* that you do. This conclusion holds for both quantitative and qualitative methods.

5.4. *Case-specific Explanation: An Unsustainable Claim*

Some embrace QCA as deterministic (e.g., Soulliere 2005; Mahoney 2008). Amenta and Poulsen (1994:24), in a special *Sociological Methods and Research* issue coedited by Charles Ragin, write

approvingly that "like Mill's methods . . . QCA is also deterministic rather than probabilistic in most applications." Ragin has vacillated about QCA (1987, 2008) and now claims that whether the world is deterministic is a side issue that may be unknowable (Ragin, this volume, p. 82). FMR claim QCA is not deterministic while claiming that we claim QCA is deterministic. For the record, we do not. Neither Ragin nor FMR acknowledge agreeing with Lucas and Szatrowski (this volume, pp. 24, 27).

Many QCA proponents now reject determinism, but some make claims that pose an equally intractable problem. Ragin claims that researchers with case-specific knowledge will produce "particularizing explanations citing case-specific features or events" (this volume, p. 82), and alleges this salvages QCA.[2]

Of course, case-specific data can be used productively in many ways, including in constructing variables (e.g., Tach and Greene 2014), coding variables (e.g., Lucas 2013), and forming cases (e.g., Guest and Lee 1984). However, a QCA researcher committed to devising case-specific explanations is committed to inferring unit-specific effects or, at least, committed to inferring for each unit i whether effects are positive, negative, or zero. Thus, Ragin's claim is a claim that QCA solves the fundamental problem of causal inference (Holland 1986).

Consider equation (2):

$$\delta_i = \left(Y_{i1}^* | X_k = 1\right) - \left(Y_{i0}^* | X_k = 0\right), \qquad (2)$$

where Y_{i1}^* is the observed outcome for case i when the value of X_k is 1 and Y_{i0}^* is the observed outcome for case i when the value of X_k is 0. Consider also $Y_{i1} | X_k = 0$ and $Y_{i0} | X_k = 1$, the value Y would obtain for cases had X equaled 1 instead of 0 and had X equaled 0 instead of 1, respectively. The causal effect of $X_k = 1$ versus $X_k = 0$ is δ_i. The fundamental problem of causal inference is that for any case i, one can observe only one of Y_1 and Y_0. Thus, at the unit level, δ_i can be identified only in the crossover experiment (Morgan and Winship 2007), because otherwise for every case, essential information needed to identify the unit-specific effect is missing.

Many analysts solve the problem by estimating average causal effects, using $\bar{Y}_1 | X_k = 1$ to estimate $Y_{i1} | X_k = 0$ for cases where $X_k = 0$, and they make a similar move to estimate $Y_{i0} | X_k = 1$. The analogue for an analyst interested in a specific causal recipe would entail comparing

across cases or instances to establish that $X_k = 1$ is part of an outcome's causal recipe. However, as I show below, devising particular case-specific explanations prohibits such an inferential move.

Imagine a social outcome with causes expressed in the equation below:

$$Y = f(\gamma X^{\alpha}), \qquad (3)$$

where f is an unspecified function, Y is a finite matrix of outcomes on N units of interest, X is a matrix of variables, γ is a matrix of coefficients, and α is a matrix of coefficients that serve as exponents; X, γ, and α are of infinite size because X contains *all* micro- and macrolevel characteristics and all combinations thereof for N units. The unconstrained nature of f coupled with α allows an infinity of parametric and nonparametric functions.

Analysts aim to partition X as follows:

$$Y = g(\gamma^* X_a^{\alpha^*}) + h(\eta X_b^{\nu}), \qquad (4)$$

where g and h are functions; γ^* is a finite matrix of coefficients; α^* is the matching, finite matrix for exponents; X_a is a finite matrix of rank (k) composed of variables of interest; X_b is an infinite matrix of all elements of X not in X_a; and η and ν are matrices of infinite size. Probabilistic models treat $h(\eta X_b^{\nu})$ as a (usually unitary) error term, often signified by ε, a finite matrix of N e_i's, escaping the need for infinite data.

For most analysts, γ^* is a matrix of average effects—in other words, they do not seek to estimate unit-specific effects. But Ragin's unit-specific effects promise creates immediate problems, because it means that instead of estimating some aggregated causal effect, γ^*, the analyst is intent on learning whether $\gamma_{ki} X_{ki}^{\alpha}{}_{ki} = 0$—that is, whether the specific effect of X_k for unit i is zero. Alas, not only does knowing $\gamma_{ki} X_{ki}^{\alpha}{}_{ki} = 0$ not imply that $\gamma_{kj} X_{kj}^{\alpha}{}_{kj} = 0$, but actually, given an interest in unit-specific effects, there is insufficient information to even determine whether $\gamma_{ki} X_{ki}^{\alpha}{}_{ki} = 0$. It should be clear that even over the *longue durée*, Ragin's claim requires the QCA researcher to access data he or she can neither obtain nor estimate—the outcome for case i were X_k equal to a value other than what it was. Within one space-time continuum, one cannot resolve this problem for observational data without assumptions that select other cases/instances for comparison. Yet, particularized

explanations prohibit this move. Thus, given Ragin's claim, although this researcher's data needs are not infinite, they may as well be, for they are just as impossible to satisfy. Hence, the task Ragin claims QCA accomplishes is impossible.

These implications have been well established in the contemporary literature on causality. It is beyond the scope of this rejoinder to delineate how comparative historical research is positioned on this issue, but suffice it to say, most such work is cross-case/cross-instance *comparative*. As shown above, to seek unit-specific effects in observational data is to abandon comparative analysis. Thus, the aims Ragin identifies cannot salvage QCA's algorithms.

Our paper contended that no method can be deterministic. Sadly, proponents of QCA who agree that QCA is not deterministic did not embrace this moment of agreement. Sadder still, some QCA proponents claim that QCA uses case-specific information to forge case-specific explanations and that this salvages QCA. But to use case-specific information in this way is to commit to finding case-specific effects and thus to run directly into the fundamental problem of causal inference. Consequently, this claim cannot validate the method.

5.5. *Simulated Data: Violating Feynman's Maxim*

This entire debate has gone backwards. At the outset, it must be righted: *No one has a burden of proving QCA does not work; QCA proponents have the burden of proving that QCA* does *work, if QCA proponents want QCA accepted as a valid method.* A key way to establish validity is to use simulated data. QCA proponents can reject such methods, but they still have a burden to meet, and it will be *very* hard to meet it without simulated data evidence.

FMR and Vaisey acknowledge that they agree with us that simulations are useful, though FMR claim we thought simulations were needed *because* QCA is deterministic. That is not our position. Olsen (this volume) replicated our work but rejects it in part because we generate the data from our own minds, making us allegedly idealists. To Olsen, appropriate analyses use real data. And, although earlier advocating the analysis of "hypothetical cases" (Ragin, this volume, p. 81), Ragin dismisses our simulations because they come from "a researcher's imagination" (Ragin, this volume, p. 89). I do not see the difference between

constructing hypothetical cases, which Ragin advises, and constructing simulated cases, which Ragin rejects.

I identify two logical problems with QCA proponents' rejection of simulation evidence. First, setting aside Ragin's self-contradiction, QCA proponents often claim that case-specific knowledge will be used in the analytic process. Recalling equations (3) and (4), an infinite number of factors are associated with any real-world case. Simulations allow analysts to limit the factors associated with a case to a finite number. Thus, QCA proponents who reject simulated data are thus implicitly maintaining that QCA will work when dealing with the infinite number of factors in a real data case but will fail when dealing with a finite number of factors (as in the simulated data case). Such reasoning undervalues the difficulty of empirical research and contradicts all humans have painstakingly learned about the challenges inherent in social research.

Second, Olsen's (this volume) argument against and Ragin's (this volume) dismissal of analyses of simulated data raise a question: What would these analysts say to a scholar whose real-world data were an exact replica of a data matrix from our paper? This is an important question because on a planet with more than 7 billion living human beings, not to mention 193 member states of the United Nations, thousands of times that many cities and towns, approximately 63,000 multinational corporations (Gabel and Bruner 2003), octillions of nonhuman animals (minimum), centillions of plants, each having an infinite number of characteristics, it is *extremely* likely that every one of our data matrices reflects real cases, a likelihood that rises further if one considers history and the googolplex entities potentially at issue. And, any data matrix, but historical data matrices especially, may have no other accompanying information, precluding data augmentation. Thus, despite our data arrays' origin, those who require analysts to use real-world data, and especially those accepting nonprobability sampling, should be assured that there is at least one, and probably many more, set(s) of real-world cases that exactly match each data array we present.

For such sets of cases, do not our findings indicate what QCA would reveal? And, as we know the causal relations in our data, does this not indicate that QCA will fail to find the vast majority of causal relations embedded in the data, even under best conditions of having the exact variables relevant to the outcome as inputs?

5.6. *Summary of General Comments*

Solid methods cannot be built on flawed foundations. QCA's asymmetric causality is illusory. Configurations are interactions. Case-specific explanations do not salvage QCA algorithms. Simulated data do not do violence to QCA. QCA proponents may stubbornly disagree, but their claims are demonstrably incorrect. If these claims constitute part of the distinctive foundation upon which QCA is meant to stand, then, alas, QCA cannot stand.

6. TWO STUDIES AND TWO EXAMPLES

In the original draft of our paper, we included an analysis of real-world data, but reviewers called for its removal. It was part of our effort to match QCA claims (here, the resistance of some to analyses of simulated data), but we were happy to remove it for two reasons. First, we realized that any test using real-world data can be dismissed regardless of its result. Unfortunately, the pro-QCA reviewer insisted we reinsert a study of real-world data. Our second reason was a fear, and what we feared is exactly what seems to be happening—QCA proponents are largely ignoring the dozens of wrong QCA conclusions across the simulations in the paper. Instead, QCA proponents have offered to the symposium "reanalyses" of three data sets—our Study 1 data, the shuttle *example*, and the race and presidential elections *example*. It should be obvious that we offered only one of these three data sets as appropriate for systematic evaluation of QCA. Indeed, in our draft manuscript, we discussed the shuttle example and the race and presidential example in the text, rather than placing the results in a table, to underscore their illustrative nature. However, during review, the pro-QCA reviewer argued we should place the shuttle example results in a table and relabel it as Study 1. We placed results in a table as requested, but we successfully resisted labeling the shuttle example Study 1 precisely because we regard the shuttle example as indicating the stakes in play, nothing more—and nothing less. Still, some QCA proponents have now treated our example analyses as if such real-world data provide an appropriate basis for an evaluative study of QCA. We contend this is incorrect.

Despite QCA proponents' tendency to dismiss our simulated data evidence, I encourage the reader seeking to answer his or her own research questions to remember that evidence and the *extremely* serious doubt

about the utility of QCA that those simulations raise. Further, I suggest that the symposium simulations of Bowers and Seawright further corroborate our conclusions.

I am tempted to ignore analyses of our examples to avoid further misdirecting attention, but were I to ignore them, multiple errors in their commission might go unrecognized, leaving the impression that QCA proponent claims are valid. However, their claims are not valid, and believing them to be valid could harm social scientific research. Thus, I address their analyses of our examples as well as their effort to replicate our studies.

6.1. *Reanalysis of Study 1 and Dismissal of Study 3*

Vaisey (this volume, p. 108) finds the correct parsimonious solution in Study 1 and the correct complex solution if he populates all cells of the data matrix. He dismisses our Study 3 findings on the basis of his Study 1 finding. He also claims we do not provide the Z variable. Producing his own Z variable, he finds he can obtain the correct parsimonious solution, but the complex solutions include the noncausal Z, as we reported.

FMR (this volume, p. 96) reanalyze our Study 1 data and report that QCA obtains the correct parsimonious results both for $Y = 1$ and $Y = 0$. They discount our complex and intermediate solutions because "the complex and intermediate solutions are not valid tests given that these are synthetic data that do not cover all possible configurations" (this volume, pp. 96–97). They then also claim that we did not provide the noncausal Z; in response, they produce their own noncausal Z, but do not convey how they produced it. They find QCA eliminates the noncausal Z. Finally, FMR report that QCA succeeded when they eliminated the overdetermined cases.

As a preliminary comment, on the noncasual Z, all I can say is, "Mea culpa." Failing to include the Z vector in the data supplied was my oversight. The paper relates our construction of Z (this volume, p. 19), but without the random component, one cannot replicate our Z vector. Further, I thought I had included the code for this analysis in one of our appendices, which would have provided enough information to perfectly construct our Z. However, checking our appendices reveals I did not include that code. Thus, I apologize; I will post on my Web site the full data matrix for Study 1 before the symposium is published, including the noncausal Z.[3]

First, and substantively, note that both FMR and Vaisey claim that QCA does not work when the data matrix is not fully populated. Vaisey fixes the problem by just populating the data matrix, and when he does so, he reports that QCA obtains the correct complex solution. Thus, FMR and Vaisey basically indicate that QCA fails when there are zero cells. Yet, QCA was historically developed to aid moderate- and small-*N* analyses, and small-*N* data are likely to have zero cells. Unfortunately, analysts working with real-world data often cannot just populate all cells of their data matrix to use QCA. Thus, Vaisey and FMR have invalidated a major justification for QCA.

Further questions arise about their replications. Neither FMR nor Vaisey commentary reports thresholds for consistency and frequency. FMR provide no documentation, but Vaisey's Stata code shows that he uses a consistency threshold of .99, which is substantially higher than ours (.8; this volume, p. 17), while our thresholds match the advice of Ragin (2008). This difference invalidates his replication of our noncausal *Z* analysis. Notably, Vaisey's complex solutions still include noncausal *Z*; it is possible his parsimonious solution would also, had he matched our consistency threshold. Thus, Vaisey does not provide a true replication study, and FMR's work is insufficiently documented for evaluation.

Finally, both Vaisey and FMR are skeptical of complex solutions, and FMR, rejecting most of our solutions as a matter of principle, do not report their complex and intermediate solutions. I do not accept the rejection of these solutions for two reasons: (1) FMR's basis for rejecting intermediate solutions is contradictory, and (2) QCA advice on the utility of different solutions prohibits universal rejection of these solutions.

First, because FMR accept simulations, their criticism seems driven by the presence of zero cells in the data matrix (i.e., limited data). However, an assertion that the presence of zero cells invalidates complex and intermediate solutions contradicts the *raison d'être* of intermediate solutions. According to Ragin (2008:131–32), the complex solution ignores regions in the variable matrix that lack data. Parsimonious and intermediate solutions impose assumptions on those regions, with intermediate solutions imposing fewer or less severe assumptions.

We presented parsimonious solutions for completeness. But we emphasized the complex solution because otherwise we could be accused of invoking problematic assumptions to drive simpler solutions away from the true causal relations. In fact, we might have inadvertently done just that. That would make our assessment of QCA unfair.

To protect against that eventuality, we highlighted complex solutions. Even so, in the vast majority of our simulations, the parsimonious solutions are also wrong.

Note, however, that FMR reject both the complex solution and the intermediate one as invalid *because* there are regions that lack data. Both intermediate and parsimonious solutions are produced by using analysts' causal expectations to code regions lacking data. As the intermediate solution is designed for the situations where the data have holes, to claim the intermediate solution is invalid *because* the data have holes is contradictory.

Finally, a second reason I reject Vaisey and FMR's rejection of the complex and intermediate solutions is that all our study of QCA indicates QCA proponents advise analysts to use the solution(s) that make the most sense given their substantive and theoretical interests. This I have always taken as a useful QCA commitment. Thus, there is no basis now for claiming an analyst with a data matrix mirroring the one used in Study 1 would *not* have an interest in the complex or intermediate solution. Hence, if analysts' interests (even partly) determine which solutions deserve focus, the complex and intermediate solutions are relevant for our assessment of QCA.

In sum, the wholesale rejection of intermediate and complex solutions contradicts other fundamental claims QCA proponents have consistently maintained. Further, Vaisey (this volume) changed important aspects of the analysis, such that it is inappropriate to regard the work as a replication. FMR do not provide sufficient documentation to evaluate their replication effort, but they regard complex and intermediate solutions as irrelevant because they reject all such solutions on what I argue are contradictory grounds. Finally, both FMR and Vaisey impose a criterion (i.e. QCA analyses cannot work when some of the logically possible permutations of the data are not observed) that necessarily repudiates a primary justification for QCA—its claim to allow rigorous formal analyses of small-*N* data. Even so, Olsen (this volume, p. 102) explicitly indicates success in replicating our findings in the review process. Given these features of the replication effort provided by pro-QCA commentators, I stand by our analyses.

6.2. *Race and Presidential Elections Example*

Ragin analyzes our race and presidential elections example and claims set-theoretic perspectives reveal "other, more striking patterns" (this

volume, p. 87). What is the striking difference between his results and our odds ratio results? While we find that "being white is positive for candidates' presidential prospects" (Lucas and Szatrowski, this volume, p. 35), Ragin finds that "being white is a very widely shared antecedent condition for success" and that "being African American is a near perfect subset of losing candidates, indicating that this outcome is very widely shared by African American candidates" (this volume, p. 88). I fail to see the difference between his conclusion and ours.

However, there is a complication. Ragin (this volume, pp. 87–88) makes comparisons that Ragin (2008:17–20) claims are unnecessary. If one makes the two comparisons that Ragin (2008) advises, as we did in the paper in demonstrating how QCA addresses asymmetric causality, one comparison will show that "being white seems to be a cause of electoral victory" (Lucas and Szatrowski, this volume, p. 34), while the other will show that "being white seems to be a cause of electoral defeat" (Lucas and Szatrowski, this volume, p. 34). How does Ragin (this volume) escape the contradictory findings of the advised method of Ragin (2008) to reach a conclusion matching our odds ratio finding? Contrary to Ragin (2008:17–20), which claims analysts need consider only three cells of a four-cell table, Ragin (this volume) brings into consideration the fourth cell, just as conventional quantitative (and most qualitative) research would. Thus, in his comment, Ragin implicitly admits one should use all table cells, thus denying the Ragin (2008) advice.

That Ragin (this volume) now agrees with almost all other analysts means that this distinctive QCA claim, which really produces serious error, will, perhaps, be set aside. However, it would have aided the continuing methodological dialogue had Ragin explicitly acknowledged that he has changed his mind and perhaps even acknowledged this as a new point of agreement between us. Sadly, that did not happen.

6.3. *Shuttle Example*

Ragin analyzes shuttle data, replicating neither of our solutions. Instead, he finds a simpler complex solution (that contradicts previous findings of engineers, the Presidential Commission, and others) and a parsimonious solution that replicates previous findings. He then observes,

> Finally, it is important to have full disclosure and avoid giving a one-sided presentation. What I am referring to is the simple fact that L&S [Lucas and

Szatrowski, this volume] failed to present the logistic regression analysis of this data set, which is surprising given that they often use this technique in their critique. I present the missing analysis, using the original uncalibrated data. . . . (Ragin, this volume, pp. 91–92)

Ragin's logistic regression model fails to find any statistically significant effects for erosion. Finding a startling contrast between his QCA and logistic regression results, Ragin suggests that confining oneself in this case to tools "from the CQNR [conventional quantitative research] toolkit, leaves engineers in the dark (and out in the cold) about the impact of low temperatures on O-ring erosion" (this volume, p. 92).

In our research, we estimated logistic regression models for erosion and nonerosion, shown in Table 1. For erosion, we find a statistically significant negative coefficient for joint temperature; for nonerosion, we find an equal-valued, opposite-signed, statistically significant coefficient. We did not present this material because we were comfortable using prior research, much of which was qualitative, to establish the current understanding of the cause of O-ring erosion in the shuttle, and the logistic regression results were consistent with earlier findings. Ragin (this volume, p. 92) claims we "often use this [logistic regression] technique"; actually we presented only one table of logistic regression coefficients in the paper, using the model in only one study out of the six in our paper and in only one of the five supplementary studies. And, of the four data sets constructed for the paper, we used a logit model in constructing only one. Yet, to Ragin, we use this technique "often" (this volume, p. 92).

Despite the views attributed to us, we were not attempting to compare QCA to statistical analytic approaches; we were attempting to compare QCA to sound analytic approaches. The latter comparison is the

Table 1. Logistic Regression Coefficients for O-ring Erosion and Nonerosion, $N = 23$

Variable	Erosion	Nonerosion
Constant	13.573 (9.908)	−13.573 (9.908)
Field PSI	.015 (.014)	−.015 (.014)
Nozzle PSI	.016 (.016)	−.016 (.016)
Joint temperature	−.250^ (.150)	.250^ (.150)

Note: Standard errors in parentheses, caret (^) marks coefficients discernibly different from 0 at or below α = .10.

germane one. The logistic regression model is one sound method, but other sound methods had come as close to establishing the cause of O-ring erosion as may be possible. In that context, the logistic regression model would have been piling on. We chose not to do that.

Ragin (this volume, pp. 89–90) writes, "To calibrate the two pressure variables as fuzzy sets, they appear to use 50 = .05, 100 = .27, and 200 = .95. To calibrate temperature as a fuzzy set they appear to use 32°F = .05, 65°F = .50, and 98°F = .95." We write, "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" (Lucas and Szatrowski, this volume, p. 15). Why does Ragin use the language of "appears" when he knows (Ragin 2008:94–95) that setting the bounds we did using the logistic distribution as he has recommended (Ragin 2008:86–97) *simultaneously* sets the values he reports? Sadly, Ragin's word choice may suggest to some that Ragin believes there was some deception or incomplete reporting in our work.

Ragin (this volume, p. 90) then states that "O-ring erosion is coded dichotomously, even though there is some variation that could have been used to create a fuzzy-set outcome." We coded erosion dichotomously because Feynman (1986:F-1) and Boisjoly (reported in Vaughan 1996, pp. 174–175) engineering theories underlie our analysis—if there is *any* erosion, the O-rings are not functioning as designed, and thus the mission may have been endangered. Thus, the relevant concern requires coding erosion dichotomously. Sadly, in the real-world data example QCA proponents desired, the primary place that clearly implicates substantive theory, Ragin (this volume) criticizes our using theory to guide the analysis.

Ragin (this volume, p. 91) reports "repeated efforts to reproduce their [our] results, using different calibrations and different specifications of the analysis." He then settles on a different calibration for temperature—that is, in place of our bounds, which derive from NASA policy on critical values for launch, a policy based on the knowledge of shuttle designers and engineers, Ragin substitutes bounds of 45 and 65, with a midpoint of 60. What engineering theory does Ragin (this volume) use to select these bounds? He reports none. He does assert that the overarching engineering concern was the danger of low-temperature

launch. However, Vaughan's (1996) comprehensive study does not support Ragin's assertion. Engineers were unsure about the cause of erosion, which is why, over time, they changed joint and nozzle pressures and tested multiple combinations. Engineers were studying erosion long before *Challenger* made its way to the launchpad; only several launch delays of the previous shuttle and *Challenger* coupled with record-setting worsening weather eventually placed STS-51-L in a potential low-temperature situation. Indeed, the late emergence of a concern with low temperature is a key part of Vaughan's (e.g., 1996:302) story. NASA had policy on both low- and high-temperature launch, and in meetings on the eve of launch, risks of warm launches were explicitly mentioned (Vaughan 1996:316). These facts and many other case-specific facts belie Ragin's assertion that the engineers' focus fell only on low temperature.

Ragin (this volume) proceeds with his calibration values and finds QCA to replicate the earlier research. Ragin reports trying several calibrations, but we do not know how many he tried before selecting the chosen set or what criteria he used to select those values. We do know, however, that an infinite number of calibration combinations exist. It is possible some will lead QCA to find the true causal factors. Yet, if one does not already know the true causal factors, how can one use knowledge of the true cause to decide which calibrations to retain?

Finally, I was troubled at Ragin's failure to replicate our results. Thus, I double-checked our data against Vaughan (1996:442–44) and found the data matches. I then double-checked Ragin's data against Vaughan and found two discrepant cases. Ragin lists 80° for STS-3 joint temperature and 67° for STS-41-G joint temperature, whereas we (and Vaughan [1996]) list 69° and 78° for those missions, respectively. I then checked the Presidential Commission report hosted on the NASA Web site (history.nasa.gov/rogersrep/genindex.htm; henceforth html report); the relevant table is located at (history.nasa.gov/rogersrep/v1ch6.htm), which lists 80° for STS-3 and 67° for STS-41-G, matching Ragin.

Construction of the html report may have introduced inaccuracies. Alternatively, the html report may reflect corrections of errors that became known after publication. To assess these hypotheses, I turned to the published 1986 hard-copy version (Rogers et al. 1986). When I checked, I found the hard-copy table matches Vaughan (1996) data used in Lucas and Szatrowski (this volume) but does not match that in Ragin (this volume) and the html report.

I then perused the various tables at issue to see if I could discern how discrepancies might have occurred. Evidence is not definitive, but it may be telling that STS-4 follows STS-3 in the table (the table lists missions by launch date, not by assigned mission number). STS-4 was the mission whose material was lost at sea, preventing collection of blow-by and erosion data. According to the 1986 hard-copy report, STS-3 launched in 69° and STS-4 launched in 80°. However, the html report lists 80° for both STS-3 and STS-4, suggesting the data for STS-4 were assigned to STS-3. Ragin (this volume, Table 4) also lists 80° for STS-3.

Further, because STS-4 was the only mission with no data for erosion and blow-by, it is the only mission for which the 1986 hard-copy report lists "Not Applicable" for erosion and blow-by. However, the html report lists "Not Applicable" for *two* missions—STS-3 and STS-4. Because the html report records "Not Applicable" for a mission for which data were apparently collected, I conclude provisionally that the html report and Ragin's analysis incorrectly assigned the temperature data for STS-4 to STS-3.

Ragin offers no citation for his data, so where he obtained them and what was listed for STS-3 is unclear. He might have obtained the data from a source I have not found. However, if Ragin obtained data from the html report, he would have had to recode the "Not Applicable" cell for STS-3 to zero for erosion (Ragin, this volume, Table 4). While this recode matches the hard-copy report (which listed no erosion or blow-by for STS-3), if Ragin recoded "Not Applicable" in this manner, he missed a chance to discover discrepancies for STS-3 in his source given case knowledge that only one mission prior to *Challenger* failed to provide erosion and blow-by data. Any source showing "Not Applicable" for erosion and blow-by for more than one mission should elicit an investigation.

The other discrepant case is STS-41-G. Rogers et al. (1986), Vaughan (1996), and Lucas and Szatrowski assign STS-41-G and STS-51-A, adjacent missions, 78° and 67°, respectively. The html report and Ragin record 67° for both. DM-6, a test between the missions, is the only test in Rogers et al. whose adjacent missions each occupy only one dataline and contain valid erosion data. Their similarity may have facilitated the error.

In sum, study of the values assigned to STS-3 and STS-41-G suggest errors in the Presidential Commission report on the NASA Web site as of March 21, 2014. I believe these discrepancies, coupled with Ragin's choice of calibration scores, explain Ragin's failure to replicate our

shuttle example findings. Notably, the effort to uncover and reconcile data discrepancies is not an example of case-oriented work. The work is simply what any careful analyst would do when faced with a failure to replicate on ostensibly the same data—interrogate the data by inspecting data sources and, if necessary, specific cases before accepting any results.

Still, the conclusion is only provisional. Perhaps a detailed search through all five volumes of the Presidential Commission report would alter various aspects of the data matrix and ultimately change the conclusion. Further, while Feynman (1986), Vaughan (1996), Boisjoly (reported in Vaughan 1996, p. 157), and Rogers et al. (1986) all claim that the cause of erosion is low temperature, in fact, it is still possible that the QCA complex solution we obtained *is* the correct one. Ask yourself this: If an engineer were told that erosion is explained by a complex solution of a three-way interaction involving low temperature, high field pressure, and high nozzle pressure and a parsimonious solution involving low temperature alone, should that engineer conclude that pressures are irrelevant? It is not so clear that he or she should. All this is to say, first, that parsimonious solutions are not necessarily better solutions and, second and most important, that real-world examples are and will always be *insufficient grounds to accept or reject a method*, because real-world data preclude the 100 percent certainty of cause needed for a clear methodological test. Thus, we sought to have no such work in our paper. Alas, with inclusion of such an example now, just as we feared, aspects of our paper *least* relevant to evaluating QCA are receiving most of the attention of QCA proponents.

6.4. *Remaining Studies*

Now seeing QCA as stochastic, QCA proponents still reanalyze only the two deterministic data simulations. Were they correct, their reanalyses could only justify QCA for deterministic processes. Studies 2, 4, 5, and 6 in the paper, and all five supplementary studies, were basically ignored. Thus, our findings on stochastic data–which all social analysts almost always have–stand uncontested. Of course, QCA proponents may elsewhere reanalyze our stochastic data. If so, we can hope those works will acknowledge agreement when it emerges and avoid impugning others' commitments, motives, and integrity. However, as the standards of the world's leading sociological methods journal greatly exceed those of many other outlets, it is possible further reanalyses will offer even more heat but very little light.

7. TURNING HEAT INTO LIGHT: A CONCLUDING NOTE

During peer review, our expressions of professional admiration for those who take risks to develop new technology were crowded out by space constraints and the need to address several critiques. Fortunately, the editor allowed us to include five additional analyses as supplements, aiding our ability to address issues reviewers raised. But, despite corroborating simulation findings in the symposium, QCA proponents remain unsatisfied.

The burden of proof, however, really rests with QCA. Usually we assume that a technology does not work prior to evidence that it does. Yet, QCA use has preceded such evidence. To paraphrase Brian Russell after the pivotal ad hoc meeting on the eve of the *Challenger* launch, we are acting as if we can use it until we prove it is *unsafe* to use, not that we *cannot* use it until we prove it is safe to use (Vaughan 1996:325). This is backward. This is dangerous.

Alas, the accumulating evidence indicates QCA is *un*safe to use. QCA proponents remain unconvinced despite simulations showing low success rates, with one finding a success rate of 0. Setting proponents and opponents aside, should substantive analysts be confident drawing conclusions from a method with documented success rates as low as 0? Should editors devote journal space to work whose findings are based on such a method? To some, such questions may seem aggressive, but such questions—does the success rate validate use?—are asked of every method. QCA must face such questions, too.

QCA failure *or* success should be a matter for sober exploration. We all have an interest in accepting only methods that work, and we all have an interest in the most expansive set of methods we can have. In a better dialogue, critical evaluators of new methods would not be deemed sarcastic for expressing respect for the difficult work accomplished to bring a method to maturity. All would engage, while accepting that many creative initiatives prove unsuccessful, that promising developments often prove a dead end. One way to accept that possibility is to recognize that it *is* a contribution to develop a line of work enough that it can be *seen* to be the dead end that it may be and, further, that critique may help reveal the path to a heretofore hidden escape from the ostensible dead end. With such a posture, any work of rigor done conscientiously is cause for joy, regardless of what it reveals.

With that as our view, we welcome critique of our work. We were saddened by the combat of the review process, and I am saddened that combativeness—questions about intent, suggestions of deception—has animated the symposium. Combativeness does not help, and may harm, the effort to excavate the social world. Life is too short, and there is too much work to do, for such distractions. Either our methods will be strong and perhaps succeed, or they will be weak and likely fail. Testing them is key to arranging the former and avoiding the latter.

Though cognizant of the possible seriousness of the implications of any findings for our efforts to study the social world, it was with a spirit of inquiry that we invited our interlocuters to dialogue. Despite what has occurred, I now invite them again to join with us in that same spirit, to figure out how the world—and our methods of studying that world— really work. Regardless of what such work reveals, and which of our cherished views are shown to be false, everyone wins if that occurs.

Notes

1. One could study each recipe or cluster them. For example, some schools may use recipes A and B, others may use A and C, others may use D, E, and F, and so on. Each combination would be a different regime, and one could study regime correlates. Finally, one could attempt to combine regimes on theoretical or empirical grounds.

2. Of Ragin's (this volume, pp. 83–84) example of 20 cases with no variation on Y, I offer three observations. First, whether some are comfortable doing something is irrelevant; the issue is whether what they do is demonstrably valid. Second, his data cannot establish that any set of alleged causes are not exactly mirrored in 200 cases where Y differs from the value it has in his hypothetical 20 cases, undermining any causal inference drawn from any method (even though case-oriented study of one or more of those cases is fine). Third, the error of Ragin's design has been shown in the qualitative literature (Geddes 1990), so my observations reflect methodological, not quantitative, reasoning.

3. My Web site URL is SamuelRoundfieldLucas.com, and the data file is accessible from the data sets link.

References

Adams, Julia. 2005. *The Familial State: Ruling Families and Merchant Capitalism in Early Modern Europe*. Ithaca, NY: Cornell University Press.

Amenta, Edwin and Jane D. Poulsen. 1994. "Where to Begin: A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis." *Sociological Methods and Research* 23:22–53.

Anderson, Elijah. 1976. *A Place on the Corner*. Chicago: University of Chicago Press.

Cheeseman, Peter. 1985. "In Defense of Probability." Pp. 1002–1009 in *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, Vol. 2. San Francisco: Morgan Kaufmann.

Elwert, Felix. 2013. "Graphical Causal Models." Pp. 245–73 in *Handbook of Causal Analysis for Social Research*, edited by Stephen L. Morgan. New York: Springer.

Feynman, Richard P. 1986. "Appendix F—Personal Observations on Reliability of the 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. Kutyna, Sally K. Ride, Robert W. Rummel, Joseph F. Sutter, Arthur B.C. Walker Jr., Albert D. Wheelen, and Charles E. Yeager. Washington, DC: Government Printing Office.

Gabel, Medard and Henry Bruner. 2003. *Globalinc: An Atlas of the Multinational Corporation*. New York: New Press.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2:131–50.

Guest, Avery M. and Barrett A. Lee. 1984. "How Urbanites Define Their Neighborhoods." *Population and Environment* 7:32–56.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 396:940–70.

Lareau, Annette. 1989. *Home Advantage: Social Class and Parental Intervention in Elementary Education*. New York: Falmer Press.

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

Lieberson, Stanley. 2004. "Comments on the Use and Utility of QCA." *Qualitative Methods* 2:13–14.

Lucas, Samuel Roundfield. 1999. *Tracking Inequality: Stratification and Mobility in American High Schools.* New York: Teachers College Press.

Lucas, Samuel Roundfield. 2008. *Theorizing Discrimination in an Era of Contested Prejudice: Discrimination in the United States*, Vol. 1. Philadelphia: Temple University Press.

Lucas, Samuel R. 2009. "Stratification Theory, Socioeconomic Background, and Educational Attainment: A Formal Analysis." *Rationality and Society* 21:459–511.

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 Roundfield. 2013. *Just Who Loses? Discrimination in the United States*, Vol. 2. Philadelphia: Temple University Press.

Lucas, Samuel R. 2014. "An Inconvenient Dataset: Bias and Inappropriate Inference with the Multilevel Model." *Quality and Quantity* 48:1616–49.

Lucas, Samuel R. and Alisa Szatrowski. 2011. "Qualitative Comparative Analysis in Critical Perspective: Simulation, Replication, and Epistemological Analyses." Unpublished manuscript, University of California–Berkeley.

Mahoney, James. 2003. "Reflections on Fuzzy-Set/QCA." *Qualitative Methods* 2:17–21.

Mahoney, James. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41:412–36.

Morgan, Stephen L. and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.

Nietzsche, Friedrich. [1873] 1976. "On Truth and Lies in a Nonmoral Sense." Pp. 42–47 in *The Portable Nietzsche*, translated by Walter Kaufmann. New York: Penguin Books.

Pearl, Judea. 2010. "The Foundations of Causal Inference." Pp. 75–149 in *Sociological Methodology*, Vol. 40, edited by Tim Futing Liao. Thousand Oaks, CA: Sage Publications.

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

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

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

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, Arther 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: Government Printing Office.

Soulliere, Danielle M. 2005. "Pathways to Attrition: A Qualitative Comparative Analysis of Justifications for Police Designations of Sexual Assault Complaints." *Qualitative Report* 10:416–38.

Szatrowski, Alisa. Forthcoming. "Organizational Form and School Success: Examining an Integrated Bureaucratic Urban High School." Paper presented at the 2014 Annual Meeting of the American Sociological Association, San Francisco.

Tach, Laura M. and Sara Sternberg Green. 2014. "'Robbing Peter to Pay Paul': Economic and Cultural Explanations for How Lower-income Families Manage Debt." *Social Problems* 61:1–21.

Vaughan, Diane. 1996. *The* Challenger *Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University of Chicago Press.

von Wright, Georg Henrik. 1955. "Ludwig Wittgenstein, A Biographical Sketch." *Philosophical Review* 64:527–45.

Whyte, William Foote. 1943. *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: University of Chicago Press.

Author Biography

The author biography can be found on page 78 of this volume.