give rise to associations. The more metaphysical the mechanism or process, the more useful it is, I contend. It is also less likely that it will be defined or used in the same way by different scholars, but this true of all concepts in social science.

Conclusion

Constructing Cause in International Relations is a radical work in its approach to cause and knowledge. I conceive of cause as artifice and something that, as a result, rarely maps neatly on to the real world, especially the political world. I understand causal claims as rhetorical rather than scientific in their justification, and alert us to new ways of evaluating such claims. I argue for the importance of context, which explains why correlations are generally woefully imperfect. I develop a novel approach to causation based on these premises that involves tacking back and forth between the general and the particular. It allows for multiple, competing narratives that can provide explanations and serve as the basis for forecasts. I understand knowledge as *phronēsis*, not merely *epistēmē*, and liken social science, not to physics, but to clinical medicine. I am not surprised that reviewers take exception to many of my arguments.

References

Lebow, Richard Ned. 2007. "What Can We Know? How Do We Know." In *Political Knowledge and Social Inquiry*. Richard Ned Lebow and Mark Lichbach, eds. (New York: Palgrave), 1–22.

Lebow, Richard Ned. 2008. *A Cultural Theory of International Relations*. Cambridge: Cambridge University Press.

Lebow, Richard Ned. 2010. *Forbidden Fruit: Counterfactuals and International Relations*. Princeton: Princeton University Press.

Lebow, Richard Ned. 2012. *The Politics and Ethics of Identity*. Cambridge: Cambridge University Press.

Lebow, Richard Ned, ed. In progress. *Max Weber and International Relations*.

Weber, Max. 1968. *Gesammelte Aufsätze zur Wissenschaftslehre*, 3rd ed. Johannes Winckelmann, ed. Tübingen: J. C. B. Mohr (Paul Siebeck).

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Guenther Ross and Claus Wittich, eds. Berkeley: University of California Press.

Exchanges: QCA and Set Theory

Mill's Methods, Induction, and Case Sensitivity in Qualitative Comparative Analysis: A Comment on Hug (2013)

Alrik Thiem
University of Geneva, Switzerland
alrik.thiem@unige.ch

After years of silence on the method of Qualitative Comparative Analysis (QCA), the journal *Political Analysis*—generally considered *the* organ of cutting-edge political methodology—lately featured an intriguing piece by one of its editorial board members.[1] In this article, Hug (2013) criticizes (1) QCA for its reliance on Mill's methods of agreement and difference, which are inappropriate in the social sciences; (2) applied researchers for having perverted the method's original objective of deductive theory testing to generate theory; and (3) the method itself for being highly sensitive to missing data and measurement error, a fact he accuses QCA's proponents of having swept under the rug. He further claims to be offering remedies to the last problem in particular and concludes that, since the discipline now seems to agree on the central message of King, Keohane and Verba (1994) that a single standard for conducting social-scientific enquiry applies, the use of QCA may soon simply fade away.[2]

While previous criticism of QCA's case sensitivity and the emergent literature on its diagnostics has taken a largely constructive direction (Eliason and Stryker 2009; Glaesser and Cooper 2014; Goldthorpe 1997; Schneider and Wagemann 2012: 284–295; Seawright 2005; Skaaning 2011; Thiem 2014a), this comment sets out to reveal Hug's account of the relation between Mill's methods and QCA as misguided, his assertion that QCA's original purpose of deductive hypothesis testing has been perverted to build theory as a misapprehension of the method's analytical core, and his critical warnings that missing data and measurement error may influence results as comforting reassurances that QCA—an expressly case-based technique—is case-sensitive indeed. I provide a reinterpretation of Hug's three points of criticism, and argue along the way that several of his results and recommendations are incorrect or inapplicable. Furthermore, I expose his "Monte Carlo simulation" as an awkward approach that masks the combinatorial logic behind QCA. I conclude that missing data and measurement error are important topics that should be analyzed more systematically in future research, but the added value of Hug's article to this endeavor does not transcend that of the generic objections of QCA's earliest critics voiced about 20 years ago.

QCA and Mill's Methods of Agreement and Difference

In the first section of his critique, Hug—under the illusion that the "logic behind this approach [QCA] relies on the methods of agreement and the [*sic!*] difference proposed by Mill"—is surprised that "many authors employing these methods refer proudly to this long tradition of thought" when Mill himself

[1] For helpful comments and suggestions, I am very grateful to Michael Baumgartner, Reto Spöhel, Ingo Rohlfing, Carsten Schneider, and the participants at the International QCA Expert Seminar in Zurich, October 2013.

[2] Although Hug only explicitly dismisses the use of QCA for inductive purposes, his unqualified espousal (2013: 257) of Clark, Gilligan, and Golder's (2006) argument that OLS models with interaction terms are superior to QCA can only be construed as an expression of general rejection.

rejected their use in the social sciences (Hug 2013: 253).[3] In this section, I will not revisit previous exchanges on the use (fulness) of Mill's methods, but instead prove that they do not underlie QCA.

In the third book of *A System of Logic*, Mill introduced his four canons of induction, among them the method of agreement and the method of difference, the latter of which he regarded as "the most perfect of the methods of experimental inquiry" (Mill 1882: 610). In a nutshell, the method of agreement aims at uncovering simple necessary causes (Ducheyne 2008: 367): the single circumstance (antecedent) in which at least two instances of some phenomenon (consequent) under investigation agree is inferred to be the cause or effect of this phenomenon (Mill 1882: 280). In contrast, the method of difference aims at uncovering simple sufficient causes (Ducheyne 2008: 369): it reasons that the one circumstance in which at least two instances differ, the first of which exhibits the phenomenon under investigation and the second of which does not, is the cause, or an indispensable part of the cause, or the effect of the phenomenon (Mill 1882: 280).

Mill (1882: 237) had of course already recognized that a consequent is usually the result of many rather than a single antecedent. For this reason, he deemed the method of agreement inconclusive in the social sciences because it is incapable of detecting plural causation: "we are already aware of how little value this method is, in cases admitting Plurality of Causes: and social phenomena are those in which the plurality prevails in the utmost possible extent" (Mill 1882: 611). The same applies to the method of difference, which Mill declared unsuitable for its "demonstrated impossibility of obtaining, in the investigations of the social science, the conditions required for the most conclusive form of inquiry by specific experience" (Mill 1882: 610). In sum, plurality of causes posed the crucial bottleneck.

Mill (1882: 311) defined causal plurality as a consequent following "with equal uniformity, any one of several antecedents, or collections of antecedents." Paraphrased in the modern language of propositional logic, causal plurality thus refers to situations in which a disjunction of conjunctions of antecedents implies a (conjunction of) consequent(s). Mill's methods had not yet been capable of identifying such elaborate structures. It was only from the 1930s onwards that electrical engineers, mathematicians and analytic philosophers began to develop the cornerstones of modern Boolean algebra on which QCA rests. In particular, Willard Quine (1952) and Edward McCluskey (1956) introduced procedures for the reduction of complex Boolean functions. The core mechanism of QCA is often referred to as the Quine-McCluskey algorithm (QMC) in recognition of their achievements.[4]

The essence of QMC consists in the elimination of circumstances that cannot act as potential causes. More precisely, it is based on the premise that if at least two instances of a phenomenon under investigation differ in a single circumstance only, that circumstance can never be a true cause of the phenomenon. The iterative application of this principle to the data results in the identification of the complete set of all minimally sufficient combinations of conditions for the outcome of interest, irrespective of their complexity or number. QCA thus does not rely on the methods of agreement and difference. Instead, "[t]he Boolean procedures of QCA break the Millian logjam" (Hicks 1994: 99).[5]

QCA and Induction

Hug (2013: 255) continues to argue that the inductive use of QCA is "explicitly against the original design of the method" as conceived by Ragin (1987, 2000, 2008) and "highly problematic" when scope conditions are not specified. Here, Hug not only distorts Ragin's representation of QCA but also seems to have misunderstood the method's primary objective. To prove this, I demonstrate how the hypothetico-deductive testing of Boolean implications that he considers the original purpose of QCA can be easily performed by simple data parsing.

Hug replicates the analysis by Grofman and Schneider (2009), who themselves use a data set from Ragin (2000: 292) containing information about the factors hypothesized to be associated with a generous welfare state (**W**) in the population of advanced industrialized democratic countries. The condition factors are the existence of strong left parties (**P**) and strong unions (**U**), a corporatist industrial system (**C**) and sociocultural homogeneity (**S**). The conservative solution is restated in expression (1), which says that the disjunction of the conjunction of strong left parties, strong unions and a corporatist industrial system, and the conjunction of strong unions, a corporatist industrial system and sociocultural homogeneity is minimally necessary and sufficient for the existence of a generous welfare state:

$$\mathbf{PUC} \lor \mathbf{UCS} \leftrightarrow \mathbf{W} \qquad (1)$$

Hug (2013: 258) stipulates that "the scope of the theory to be inductively derived corresponds to the set of sixteen cases." One reason for a misspecification of this scope could be that the researcher has excluded a case inadvertently in preparing the data, an unlikely situation with a target population of 16 cases. More realistically, missing data on one or more variables could have led to the deletion of cases. But why should we dispense with QCA for purposes of theory-building because of the risks of missing data? To keep the argument clean, I address the issue of missing data and measurement error separately below, and focus on the broader dismissal of inductive uses of QCA in this section.

[3] Hug merely cites a single author, but Rihoux (2006) in fact talks about Mill's indirect method of difference in relation to Przeworski and Teune's (1970) most-similar-systems-design. What is more, Mill and many of his contemporaries rejected the use of *all* experimental methods in the social sciences, including the method of concomitant variation, a forerunner of the modern correlation coefficient (Brown 1997; Mill 1882: 608–613; Rodgers and Nicewander 1988: 60), which Hug himself has used in his work.

[4] Note that not all of QMC's components are appropriate for purposes of causal data analysis with QCA, and that algorithms other than QMC are in use nowadays (Thiem and Dușa (2013a), Thiem 2014b).

[5] Schneider and Wagemann (2012: 9) also point out that Mill's methods do not require truth tables—a central feature of QCA.

Hug (2013: 252) contends that "only in a very few research areas are our theories sufficiently advanced to yield deterministic hypotheses," but if the overriding goal of science resides in the advancement of our theories from the *best* currently available (in the Popperian sense of *surviving*) to successively better ones, then I wonder why Hug so firmly rejects the potential contributions of inductive research to this effort. After all, it should be irrelevant *how* we get to better theories (ethics aside), as long as we get there *at all*. The recipe for scientific progress consists in refining methods, revising scope conditions and reducing bias (Boswell and Brown 1999: 158), but certainly not in condemning inductive or deductive research. As Goldstone (2004: 39) rightly reminds sticklers for either paradigm, "[g]reat scholars often work both sides of the fence."[6]

More concretely, let me illustrate why Hug's reference to QCA is ill-placed if he only believes in the hypothetico-deductive approach. Suppose a researcher entertained the hypothesis (H_1) that the absence of a strong left party or the absence of sociocultural homogeneity was a necessary and sufficient condition for the absence of a strong welfare state:

$$\neg\mathbf{P} \lor \neg\mathbf{S} \leftrightarrow \neg\mathbf{W} \qquad (H_1)$$

By replacing equivalence and implication operators (H_1) can be reformulated and expanded as shown in expression set (2a) to (2g).[7]

$$
\begin{aligned}
\neg\mathbf{P} \lor \neg\mathbf{S} \leftrightarrow \neg\mathbf{W} &\equiv (\neg\mathbf{P} \lor \neg\mathbf{S} \rightarrow \neg\mathbf{W}) \land (\neg\mathbf{W} \rightarrow \neg\mathbf{P} \lor \neg\mathbf{S}) && \text{(2a)}\\
&\equiv [\neg(\neg\mathbf{P} \lor \neg\mathbf{S}) \lor \neg\mathbf{W}] \land [\neg(\neg\mathbf{W}) \lor (\neg\mathbf{P} \lor \neg\mathbf{S})] && \text{(2b)}\\
&\equiv [(\mathbf{P} \land \mathbf{S}) \lor \neg\mathbf{W}] \land [(\neg\mathbf{P} \lor \neg\mathbf{S}) \lor \mathbf{W}] && \text{(2c)}\\
&\equiv (\mathbf{P} \lor \neg\mathbf{W}) \land (\mathbf{S} \lor \neg\mathbf{W}) \land (\neg\mathbf{P} \lor \neg\mathbf{S} \lor \mathbf{W}) && \text{(2d)}\\
&\equiv (\mathbf{P} \land \mathbf{S} \land \mathbf{W}) \lor (\neg\mathbf{P} \land \mathbf{S} \land \neg\mathbf{W}) \lor (\neg\mathbf{P} \land \neg\mathbf{S} \land \neg\mathbf{W}) && \text{(2e)}\\
&\qquad \lor (\neg\mathbf{S} \land \mathbf{P} \land \neg\mathbf{W}) \lor (\neg\mathbf{S} \land \neg\mathbf{P} \land \neg\mathbf{W}) &&\\
&\equiv (\mathbf{P} \land \mathbf{S} \land \mathbf{W}) \lor [(\neg\mathbf{P} \land \neg\mathbf{W}) \land (\neg\mathbf{S} \lor \mathbf{S})] \lor [(\neg\mathbf{S} \land \neg\mathbf{W}) \land (\neg\mathbf{P} \lor \mathbf{P})] && \text{(2f)}\\
&\equiv (\mathbf{P} \land \mathbf{S} \land \mathbf{W}) \lor (\neg\mathbf{P} \land \neg\mathbf{W}) \lor (\neg\mathbf{S} \land \neg\mathbf{W}) && \text{(2g)}
\end{aligned}
$$

Expression (2g) shows most clearly that a test of (H_1) only requires the identification of cases for which **P** and **S** occur in conjunction with **¬W** or, alternatively, **¬P** or **¬S** together with **W**.[8] Both Ireland and Belgium thus falsify (H_1) in Hug's data, but neither truth tables nor Boolean minimization—the two defining features of QCA—are needed to perform this test in the best Popperian manner.[9] In summary, the primary objective of QCA does not consist in deductive hypothesis testing as Hug claims, but the *identification of minimally necessary disjunctions of minimally sufficient conjunctions of conditions with respect to one or several outcomes*. The supplementation of this end with single or multiple phases of inference-statistical testing for isolated implication or equivalence hypotheses such as those proposed by Braumoeller and Goertz (2000), which Hug appears to consider *the* solution to QCA's putative problems, or those introduced by Ragin (2000), Eliason and Stryker (2009) or Longest and Vaisey (2008), is a different matter unrelated to the central mechanisms and features of this method.

Missing Data and Measurement Error in QCA

Hug conducts four "Monte Carlo simulations" in order to test how often the true model fails to materialize. In the first simulation, single cases are dropped from the data; in the second, pairs of cases are dropped; in the third, the outcome for single cases is recoded from 1 to 0 and *vice versa*; and in the fourth simulation, the outcomes for pairs of cases are recoded from 1 to 0 and *vice versa*. The first set of simulations thus deals with sensitivity to missing data, the second with sensitivity to selective measurement error in the outcome factor.

Sensitivity to Missing Data

In order to demonstrate QCA's sensitivity to missing data, Hug pulls out the sledgehammer of what he calls "Monte Carlo simulation" to crack the nut of computing the probability of retaining the true solution in expression (1) when single cases are dropped. Irrespective of the fact that "Monte Carlo simulations" are methods for the production of random variables to approximate univariate and multi-dimensional integrals (Robert and Casella 2010) whereas Hug (2013: 259) runs a simple loop over the data, he finds this probability to be 0.875, while **PUC** and **UCS** each appear as the other solutions with probability 0.0625—changes he deems "far from innocuous." To any informed methodologist of QCA, this conclusion and the method whereby it has been arrived at would be disappointing. Can we do better in arriving at the retention probability of the original solution and in assessing solution changes?

To begin with, nine out of all 16 cases lack a generous welfare state. Dropping any one of them will never change the true solution if logical remainders are excluded. Furthermore,

[6] A prime example of how deduction and induction can be melded to help improve theory is Brown and Boswell (1995), who combine game theory and QCA.

[7] I provide all steps for easier traceability of the reformulation.

[8] This can also be shown by perfect induction in a truth table, where the configurations {**P** = 1, **S** = 1, **W** = 0}, {**P** = 0, **W** = 1} and {**S** = 0, **W** = 1} are the only three for which (2g) is false.

[9] The invocation of the Boolean laws of distribution and complement in going from (2e) to (2g) may be considered an application of Boolean minimization, but QCA's procedures are considerably more demanding.

Austria, Denmark, Finland, Norway and Sweden all belong to the same positive configuration so that dropping single cases from this set could never change the true solution, either. Only if Belgium or Ireland was excluded would the solution change. No truth tables, no Boolean minimization and no simulation are required to gain this insight, just simple consideration of the data. Instead of analyzing crucial sets of cases representing positive configurations, however, Hug runs the whole gamut of QCA for all scenarios. Moreover, he fails to note that **PUC** and **UCS** are subsets of the original solution, both of which describe five out of six cases simultaneously. Put differently, instead of primarily inducing different solutions, a disambiguation of the original solution has in fact been caused by dropping single cases.

In the second simulation, Hug (2013: 259) increases the complexity of the analysis by dropping pairs of cases and concludes that "[w]hen two observations are dropped, the share of Monte Carlo runs yielding different solutions that [*sic!*] increases considerably." Out of all 120 possible pairs, the true solution fails to materialize with probability 0.24. Again, simple combinatorics would have sufficed, but Hug generally fails to understand that QCA solutions will only change if the number of cases to be dropped is at least as large as at least one positive configuration of cases. Let N be the total number of cases, c_i the number of cases in configuration i, D the number of cases to be dropped and C the number of configurations for which $c_i \leq D$. Then, equation (3) returns the exact retention probability $P_\Theta(D)$ of the original solution, for any number of dropped cases D up to N.

$$P_\Theta(D) = 1 - \left(\left(\sum_{i=1}^{C} \binom{N-c_i}{D-c_i} - E\right) \Big/ \binom{N}{D}\right) \quad (3)$$

The term E is an adjustment factor that corrects for double counting in a union of sets with positive intersections (Hohn 1966: 261ff.).[10] For example, if $D = 2$ as in Hug's second simulation, $P_\Theta(2)$ is given by equation (4).

$$P_\Theta(2) = 1 - \left(\left(\binom{15}{1} + \binom{15}{1} - 1\right) \Big/ \binom{16}{2}\right) \approx 0.7583. \quad (4)$$

More in the abstract, it is not surprising that dropping an instance of a unique configuration affects the final solution in this data set under the given parameters. It is even less surprising that dropped pairs of cases affect solutions to an even larger extent. And if Hug had carried on to drop all 65382 combinations of cases that could have been formed from 3 to 14-way combinations with the eventual finding that P_Θ declines further from 0.65, over 0.55, 0.46, 0.37, 0.30, etc. to finally 0, it would not have been distressing, either. On the contrary, I would have taken considerable comfort from these figures. It might strike methodologists like Hug, who have been trained to think of case sensitivity as something inherently negative, as incomprehensible but what if not case sensitivity is desired if analysts set up a case-sensitive research design and employ an expressly case-based method?

Sensitivity to Measurement Error

The third and fourth simulations deal with measurement error in the outcome factor, but only with respect to single cases and pairs of cases. Just as in the preceding section, I first briefly demonstrate why simulation is unnecessary because even simpler combinatorics than those used before would have yielded all relevant findings. In addition, however, for this set of simulations I also prove Hug's results to be incorrect, insofar as the extent of sensitivity to measurement error, which he regards as highly worrisome, is even massively understated.

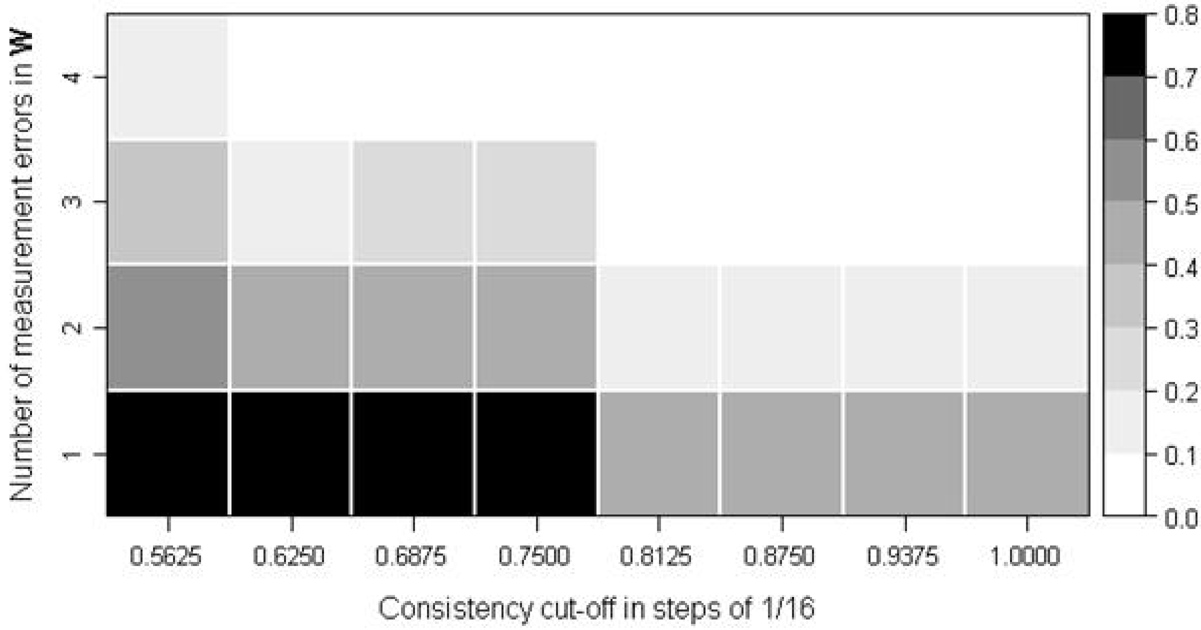
As in the case of missing data, for Hug (2013: 256), "a single case may lead us to reject a deterministic hypothesis […]", so "we should worry, it seems, considerably about possible errors." Proceeding from the assumption of determinism, Hug's results from the third and fourth simulation are incorrect. Most importantly, he fails to report **pUCS** ∨ **PUCs**, as a result of which the correct retention probability is only 0.4375, not 0.75, while **pUCS**∨ **PUCs** has a probability of 0.3125. The reason behind Hug's error is obvious. Against his assumption of determinism and due to a failure to care about consistency, disconfirming evidence was allowed to enter the minimization process.[11] For the first twelve recodes of the value on the outcome to have the described but false effect, a probabilistic subset relation with a consistency of 0.8 between the configurations and the outcome would have had to be accepted. Then, the recoding of any of the positive cases apart from Belgium and Ireland would have been ineffective. However, the correct figures could have easily been obtained without the exercise of exhaustive brute-force loops. Australia, Canada, France and the USA exhibit the same negative configuration, as do Germany, the Netherlands and Switzerland. Recoding any one of these seven cases does not change the true solution. Put simply, 7/16 = 0.4375.

In the fourth simulation, Hug again ratchets up the complexity of the analysis by recoding two values in each run. Again, under the assumption of determinism he entertains about QCA, the results are incorrect. **PUCs** and **pUCS** (both with probability 0.042) are missed and incorrect figures for **PUC** and **UCS** (both 0.058) are provided in consequence. But just as before, arriving at the retention probability of the true solution with combinatorics is almost trivial. Only recoding

[10] An alternative yet much simpler formula is $P_\Theta = \prod_{i=1}^{C}(1 - p^{c_i})$, which assumes independence in the missing-data generating process rather than dependence as Hug implicitly does. Assuming dependence leads to massive computational problems with larger number of positive configurations because the complexity of E in equation (3) increases geometrically.

[11] Hug seems to be unaware that including contradictions in the minimization implies a non-unity consistency cut-off (the measure of consistency was not yet included in the outdated *QCA* package of version 0.1-3 that he used (Dușa and Thiem 2014; Thiem and Dușa 2013b)). This is evident from footnote 25, page 259, where he claims that "all causal paths presented in a basic QCA have a perfect consistency […]." Obviously, in the presence of measurement error, initially consistent configurations can become inconsistent and *vice versa*.

Figure 1: Retention Probability of PUC ∨ UCS as a Function of Measurement Error and Consistency



pairs of negative cases that are instances of non-unique configurations will leave the original solution unchanged. Since there are seven such cases, $\binom{7}{2} = 21$ of these pairs exist. Unsurprisingly, Hug's simulation yields 21 scenarios in which the true solution is retained.

In contrast to missing data, measurement error influences the results in Hug's example in interaction with consistency, a particularity he misses in addition. Generally, the higher the consistency score, the lower the retention probability. Figure 1 depicts this probability as a function of the number of measurement errors, ranging from one to four, and the consistency cut-off, running from 0.5625 to 1.0 in steps of 1/16.[12] The crucial consistency fault line is between 0.75 and 0.8125. Also note that decreasing consistency cut-offs do not necessarily entail higher retention probabilities given the same number of measurement errors. For instance, at a consistency cut-off of 0.75 and three errors it is about 0.05 higher than for a cut-off of 0.625.

Conclusion

Mahoney (2004: 21) has provided a list of best practices for how to respond to various types of QCA critics, including *uninformed dismissers*, *informed skeptics,* and *critical innovators*. The first group includes scholars who "may believe that fs/QCA [*or QCA in general*] does not offer techniques that differ from mainstream statistical analysis," the second group those "who are knowledgeable about fs/QCA practices, but who nevertheless hold reservations," and the third group those "who identify important problems with fs/QCA but are still motivated to suggest partial or full solutions for these problems."[13]

In this comment, I have responded to Hug's criticism that QCA is based on Mill's methods, that inductive use goes squarely against its purpose, and that results are sensitive to missing data and measurement error. The reinterpretation of these points has demonstrated that QCA does not derive from Mill's methods of agreement and difference, that deductive hypothesis testing is not its original goal, and that effects from missing data and measurement error are not only to be expected but even considered useful when researchers employ case-based methods as part of case-sensitive research designs. I have shown how combinatorics can help analyze such effects more efficiently and faithfully to QCA's logic. Furthermore, I have extended Hug's elementary design to scenarios involving interactions between measurement error and consistency cut-offs. Moreover, mistakes in his original analysis have been corrected.

I do not arrogate the right to place Hug within the taxonomy of critics proposed by Mahoney. However, while his contribution is to be welcomed insofar as it directs attention to issues that should receive more notice in future research, I am left astonished by his conception of QCA as little more than another version of Mill's methods, his belief that deductive hypothesis testing is its primary purpose, the superficial and decontextualized interpretation of the effects of missing data and measurement error, and the disproportionate accusations he levels at the method's developers and users. With an unshakable conviction that QCA will not now begin to fade away when it has not done so during the 20 years since King, Keohane and Verba (1994), I hope to have illustrated in this

[12] Beyond four errors, the retention probability drops below 0.1 for all consistency cut-offs.

[13] Italicized text in square brackets added. Note that Mahoney refers to the method of fsQCA, not the eponymous software.

comment how critical innovators (as one of whom I consider myself) might contribute in a more constructive manner to future research on all aspects of this still relatively novel and exciting technique.

References

Boswell, Terry and Cliff Brown. 1999. "The Scope of General Theory: Methods for Linking Deductive and Inductive Comparative History." *Sociological Methods & Research* 28 (2): 154–185.

Braumoeller, Bear F., and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44 (4): 844–858.

Brown, Cliff and Terry Boswell. 1995. "Strikebreaking or Solidarity in the Great Steel Strike of 1919: A Split Labor Market, Game-Theoretic, and QCA Analysis." *American Journal of Sociology* 100 (6): 1479–1519.

Brown, Robert. 1997. "Artificial Experiments on Society: Comte, G.C. Lewis and Mill." *Journal of Historical Sociology* 10 (1): 74–97.

Clark, William Roberts, Michael J. Gilligan, and Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14 (3): 311–331.

Ducheyne, Steffen. 2008. "J.S. Mill's Canons of Induction: From True Causes to Provisional Ones." *History and Philosophy of Logic* 29 (4): 361–376.

Duşa, Adrian, and Alrik Thiem. 2014. *QCA: A Package for Qualitative Comparative Analysis, R Package Version 1.1.-3*. URL: http://cran.r-project.org/package=QCA/.

Eliason, Scott R. and Robin Stryker. 2009. "Goodness-of-Fit Tests and Descriptive Measures in Fuzzy-Set Analysis." *Sociological Methods & Research* 38 (1): 102–146.

Glaesser, Judith and Barry Cooper. 2014. "Exploring the Consequences of a Recalibration of Causal Conditions when Assessing Sufficiency with Fuzzy Set QCA." *International Journal of Social Research Methodology* 17 (4): 387–401.

Goldstone, Jack A. 2004. "Response: Reasoning about History, Sociologically…" *Sociological Methodology* 34 (1): 35–61.

Goldthorpe, John H. 1997. "Current Issues in Comparative Macrosociology: A Debate on Methodological Issues." *Comparative Social Research* 16: 1–26.

Grofman, Bernard and Carsten Q. Schneider. 2009. "An Introduction to Crisp Set QCA, with a Comparison to Binary Logistic Regression." *Political Research Quarterly* 62 (4): 662–672.

Hicks, Alexander. 1994. "Qualitative Comparative Analysis and Analytical Induction." *Sociological Methods & Research* 23 (1): 86–113.

Hohn, Franz E. 1966. *Applied Boolean Algebra: An Elementary Introduction*. 2nd ed. New York: Macmillan.

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

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

Longest, Kyle C. and Stephen Vaisey. 2008. "fuzzy: A Program for Performing Qualitative Comparative Analyses (QCA) in Stata." *Stata Journal* 8 (1): 79–104.

Mahoney, James. 2004. "Reflections on Fuzzy-Set/QCA." *Qualitative Methods* 2 (2):17–21.

McCluskey, Edward J. 1956. "Minimization of Boolean Functions." *Bell Systems Technical Journal* 35 (6): 1417–1444.

Mill, John Stuart. 1882. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. 8th ed. New York: Harper & Brothers.

Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience.

Quine, Willard V. 1952. "The Problem of Simplifying Truth Functions." American Mathematical Monthly 59 (8): 521–531.

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.

Rihoux, Benoît. 2006. "Qualitative Comparative Analysis (QCA) and Related Systematic Comparative Methods: Recent Advances and Remaining Challenges for Social Science Research." *International Sociology* 21 (5): 679–706.

Robert, Christian P. and George Casella. 2010. *Introducing Monte Carlo Methods with R*. New York: Springer.

Rodgers, Joseph L. and W. Alan Nicewander. 1988. "Thirteen Ways to Look at the Correlation Coefficient." *American Statistician* 42 (1): 59–66.

Schneider, Carsten Q. and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis (QCA)*. Cambridge: Cambridge University Press.

Seawright, Jason. 2005. "Qualitative Comparative Analysis vis-à-vis Regression." *Studies in Comparative International Development* 40 (1): 3–26.

Skaaning, Svend-Erik. 2011. "Assessing the Robustness of Crisp-Set and Fuzzy-Set QCA Results." *Sociological Methods & Research* 40 (2): 391–408.

Thiem, Alrik. 2014a. "Membership Function Sensitivity of Descriptive Statistics in Fuzzy-Set Relations." *International Journal of Social Research Methodology* 17 (6): 625–642.

Thiem, Alrik. 2014b. "Navigating the Complexities of Qualitative Comparative Analysis: Case Numbers, Necessity Relations, and Model Ambiguities." *Evaluation Review* 38 (6): 487–513.

Thiem, Alrik and Adrian Duşa. 2013a. "Boolean Minimization in Social Science Research: A Review of Current Software for Qualitative Comparative Analysis (QCA)." *Social Science Computer Review* 31 (4): 505–521.

Thiem, Alrik and Adrian Duşa. 2013b. "QCA: A Package for Qualitative Comparative Analysis." *The R Journal* 5 (1): 87–97.

We Need an Open Discussion of QCA's Limitations: A Comment on Thiem

Simon Hug

University of Geneva, Switzerland

simon.hug@unige.ch

My article on some limitations of Qualitative Comparative Analyses (QCA) in *Political Analysis* (Hug 2013) aims at warning scholars who use this method about its sensitivity to measurement error and problems that arise when employing it for inductive purposes.[1]

[1] Greatly appreciated comments on an earlier draft of this response were provided by David Collier, Katherine Michel, Jack Paine, and

It is heartening to note that Alrik Thiem (2014) in his response to my article agrees with both of these points. Thus he notes that "it is not surprising that the dropping of an instance of a unique configuration affects the final solution in this data set under the given parameters" and that QCA displays a considerable "extent of sensitivity to measurement error" (22).

Nevertheless, Thiem considers my critique based on these two limitations to be "misguided," (2014: 19) and also suggests that the link I draw between Mill's (1843/1973) well-known methods and QCA, used to motivate my contribution, is equally misguided. In addition, while I argued in my article that QCA might have its use to assess well-specified claims of deterministic hypotheses, either in the absence of measurement error or when such measurement errors are explicitly addressed, Thiem concludes that I am an opponent of QCA. This conclusion is based on my suggestion to rely on Braumoeller and Goertz's (2000) method to test necessary conditions in the presence of measurement error, and due to my favorable reference to the article by Clark, Gilligan and Golder (2006).[2]

As I believe the position towards QCA advocated in my article is sufficiently clear—readers should note the much more fundamental critiques by other authors[3]—I will focus on my three purportedly misguided criticisms of QCA, mostly dealing with the two that were the main object of my article: measurement error and the inductive use of QCA. I will start with a short discussion of the link between Mill's (1843/1973) methods and QCA, before moving to the two other criticisms.

Mill's Methods and QCA

My discussion of Mill's (1843/1973) methods was mostly intended as a motivation for addressing the limitations of QCA. Thiem (2014: 20) seeks to "prove that [these methods] do not underlie QCA." More specifically, he argues that QCA instead addresses one of the problems Mill (1843/1973) identified for his methods of agreement and difference—namely the plurality of causes—by allowing for several causal paths. Interestingly, prominent QCA scholars see a much more intimate link: "Qualitative Comparative Analysis (QCA) (Ragin 1987, 2000) can be seen as an extension of John Stuart Mill's well-known methods into a systematic (computer-based) comparative approach" (Schneider and Wagemann 2006: 752). Similarly, Coverdill and Finlay (1995) discuss these intimate links, which are equally present in Ragin's (1987: 36–42) presentation of his method. Ragin states himself that Mill's methods: "form the core of the case-oriented strategy" (42). While QCA may address the limitation of Mill's methods in accounting for plural causes, other limitations apply equally to QCA (see Lieberson 1991: 318, n. 11; Bennett 2004: 32).

Deductive Use of QCA

Based on a series of references to Ragin (1987), I argued in my article that the primary purpose of QCA is to test theoretical propositions in a deductive manner. This primary purpose has become diluted, however, and more and more scholars use QCA to generate theoretical propositions. I argued that this is problematic in the absence of clear scope conditions, which are hard to define for an inductively generated theory.[4]

Thiem argues that QCA is not designed to test theories deductively. More precisely, he seems to argue that the deductive part of QCA consists only of identifying the factors that are "hypothesized to be associated with" a possible outcome. Based on these factors, if a researcher were to have clear expectations about what configurations of these factors form necessary or sufficient conditions for the outcome, then QCA's task "can be easily performed by simple data parsing" (2014: 20). Hence, the algorithm underlying QCA—i.e., the Quine-McCluskey algorithm—would no longer be required.

This seems like an odd justification for a method. The parallel in the statistical world would be not to use maximum-likelihood estimation, which in all statistical packages relies on maximization algorithms, if the likelihood function to be maximized allows for an analytical solution. As Collier (2014) nicely shows, this seems to put an algorithm before the issue of how to proceed to causal inferences. Relatedly, Paine (forthcoming) carefully argues that if a researcher has clear expectations about what combinations of factors form a necessary or sufficient condition for an outcome, then evaluating this hypothesis using a regression framework will yield the same conclusions as QCA.

In addition, Thiem suggests that in the absence of theories allowing us to form deterministic hypotheses, we should draw on QCA to improve our theories. As I tried to show in my article without clear scope conditions this is a risky strategy.[5] A main problem with proceeding inductively, as identified by Paine (forthcoming), is that an empirical pattern consistent with a deterministic hypothesis is often likely to appear even if the underlying data-generating process is probabilistic. Consequently, it seems a risky path to use QCA inductively, as suggested by Thiem, to arrive at theories with deterministic elements. A useful path to protect oneself against these risks might be to implement split sample designs (e.g., Angrist and Krueger 1995), i.e. to use part of a sample of cases to inductively generate empirical generalizations, which are then deductively evaluated either by a QCA or the way Paine (forthcoming) proposes for evaluating deterministic conditions.[6] Nevertheless, it seems odd to justify the purpose of a method simply on the basis of the algorithm underlying it.

Jason Seawright.

[2] By this same standard I would have to be considered an opponent of non-linear models used in quantitative research as in a paper written almost at the same time as the one on QCA. I highlighted some inferential problems these models faced in the presence of measurement error in the dependent variable (Hug 2010). Yet I am certainly not an opponent of these models.

[3] Seawright 2013, Collier 2014, Krogslund, Choi and Poertner (forthcoming), Krogslund and Michel 2014b, Lucas and Szatrowski 2014, Michel and Krogslund 2014, Tanner 2014.

[4] For an insightful discussion, see Goertz and Mahoney (2009: 313–315).

[5] For the same point, see again Goertz and Mahoney (2009: 313–315).

[6] Proceeding in this way is also envisioned by Ragin and Rihoux (2004: 23) in reaction to a point raised by Lieberson (2004).

Measurement Error and Missing Observations

Regarding measurement error, it bears noting that my article on QCA was written in parallel to my related article on measurement error in dependent variables when using non-linear probit models (Hug 2010). To demonstrate inferential problems caused by such error, I employed commonly used Monte Carlo simulations. I relied on the same general approach to assess the consequences of measurement error and missing observations, as I also did in my QCA article.

Thiem criticizes my use of Monte Carlo on two fronts. First, he sees these simulations as a "sledgehammer" (2014: 21) and inappropriate; second, he argues that my results could also have been obtained analytically.

It is unclear from Thiem's (2014) argument why Monte Carlo simulations appear inappropriate to him. He refers to one branch of statistical methods where these simulations are actually used "for the production of random variables to approximate univariate and multi-dimensional integrals" (2014: 21). However, the origins of Monte Carlo go much further back and were prominently used to introduce stochastic elements in simulation models (see for instance Forrester 1969). In the statistical literature Monte Carlo simulations are equally used to assess (often small sample) properties of estimators with the help of simulated data.[7] Interestingly, even Ragin and Rihoux (2004: 23) implicitly envision drawing random subsamples from an initial set of cases and subjecting these subsamples to a QCA, i.e., in effect starting with a Monte Carlo randomization.[8]

Thus, Monte Carlo simulations provide a useful tool in two significant regards. The simulations help assess the properties of statistical estimators (as well as assessing any method of causal inference, see Seawright 2013); and they can help detect shortcomings of QCA, as demonstrated in the recent studies by Krogslund, Choi, and Poertner (forthcoming), Krogslund and Michel (2014a, 2014b), Michel and Krogslund (2014), and Seawright (2014). While these studies rely in part on a known data-generating process to create artificial datasets, I used an existing dataset and assumed different degrees of measurement error and missing data. As I discussed in my paper, randomly assigning measurement error in the outcome for a small set of cases is largely equivalent to considering each possible change as due to measurement error (and the same thing holds for missing data).

Consequently, I have no hesitation to admit that the same results I presented in my paper could also have been arrived at analytically. But this does not invalidate my conclusion that QCA is sensitive to measurement error in the outcome variable and to missing data, whether a "sledgehammer" (i.e., Monte Carlo simulations) or analytical tools are used.

Conclusion

My article is part of a larger set of contributions demonstrating some limitations and weaknesses of QCA. Seawright (2014) shows that QCA is vulnerable to false positives, even with a deterministic data-generating process. Krogslund, Choi and Poertner (forthcoming) note the sensitivity of fuzzy-set QCA to calibration decisions. Michel and Krogslund (2014), extending Seawright's work (2013, 2014), highlight that even with known data-generating processes, QCA has difficulties in recovering the true causal paths; in a subsequent paper these authors assess an alternative to the Boolean-algorithm used in QCA (Krogslund and Michel 2014a). Finally, Seawright (2014) addresses a limitation that Bennett (2004: 32) already considered for Mill's (1843/1973) methods, namely limited diversity.

All these contributions carry out tasks that are the daily bread of scholars interested in quantitative methods: identifying the limitations of a given method, cautioning scholars using the method, and hopefully proposing ways to address the limitations. These efforts have reinforced the strength of various methods.

Surprisingly, some scholars wedded to QCA appear not to be interested in a discussion of the limitations of their preferred method, as the tone of their responses sometimes suggests (see Ragin 2005 on Seawright 2005, Ragin 2014 on Lucas and Szatrowski 2014; and Thiem 2014). It is equally interesting that an early article highlighting issues of measurement error for QCA (Coverdill and Finlay 1995) is barely cited and referenced only once by a scholar who has written extensively on QCA—Rihoux (2003).

It is obviously easier to imply, as does Thiem (based on Mahoney 2004), that authors discussing limitations of QCA have not understood this method.

[7] For an older and a more recent example see Hartman (1991); and Aldrich, Montgomery, and Sparks (2014).

[8] I was alerted to this point by Seawright (2013: 13).

References

Aldrich, John H., Jacob M. Montgomery and David B. Sparks. 2014. "Polarization and Ideology: Partisan Sources of Low Dimensionality in Scaled Roll Call Analyses." *Political Analysis* 22(4): 435–456.

Angrist, Joshua D. and Alan B. Krueger. 1995. "Split-Sample Instrumental Variables Estimates of the Return to Schooling." *Journal of Business & Economic Statistics* 13(2): 225–235.

Bennett, Andrew. 2004. "Case Study Methods: Design Use and Comparative Advantages." In *Models, Numbers, and Cases. Methods for Studying International Relations*. Detlef F. Sprinz and Yael Wolinsky-Nahmias, eds. (Ann Arbor: University of Michigan Press), 19–55.

Braumoeller, Bear F. and Gary D. Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44(4): 844–858.

Clark, William Roberts, Michael J. Gilligan and Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses." *Political Analysis* 14(3): 311–331.

Collier, David. 2014. "Comment: QCA Should Set Aside the Algorithms." *Sociological Methodology* 44(1): 122–126.

Coverdill, James E. and William Finlay. 1995. "Understanding Mills Via Mill-Type Methods: An Application of Qualitative Comparative Analysis to a Study of Labor Management in Southern Textile Manufacturing." *Qualitative Sociology* 18(4): 457–478.

Goertz, Gary and James Mahoney. 2009. "Scope in Case Study Research." In Sage Handbook of Case-Based Methods. David Byrne and Charles C. Ragin, eds. (Thousand Oaks, CA: Sage), 307–317.

Hartman, Raymond. 1991. "A Monte Carlo Analysis of Alternative Estimators in Models Involving Selectivity." *Journal of Business & Economic Statistics* 9(1): 41–49.

Hug, Simon. 2010. "The Effect of Misclassifications in Probit Models. Monte Carlo Simulations and Applications." *Political Analysis* 18(1): 78–102.

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

Krogslund, Chris, Donghyun Danny Choi, and Mathias Poertner. Forthcoming. "Fuzzy Sets on Shaky Ground: Parametric and Specification Sensitivity in fsQCA." *Political Analysis*.

Krogslund, Chris and Katherine E. Michel. 2014a. "Can QCA Do Causal Inference? An Assessment and Proposed Alternative." Paper prepared for presentation at the Annual Meeting of the American Political Science Association, Washington DC.

Krogslund, Chris and Katherin E. Michel. 2014b. "A Larger-N, Fewer Variables Problem? The Counterintuitive Sensitivity of QCA." *Qualitative & Multi-Method Research* 12(1): 25–33.

Forrester, Jay Wright. 1969. *Principles of Systems*. Cambridge, MA: Wright-Allen Press.

Lieberson, 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(2): 307–320.

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

Lucas, Samuel R. and Alisa Szatrowski. 2014. "Qualitative Comparative Analysis in Critical Perspective." *Sociological Methodology* 44(1): 1–79.

Mahoney, James. 2004. "Reflections on Fuzzy-Set/QCA." *Qualitative Methods* 2(2): 17–21.

Michel, Katherine E. and Chris Krogslund. 2014. "Testing the Ability of Set-Theoretic Comparative Methods to Recover Data Generating Processes." Paper prepared for presentation at the Annual Meeting of the Midwest Political Science Association, Chicago.

Mill, John Stuart. 1843/1973. *Of the Four Methods of Experimental Inquiry. In A System of Logic, Ratiocinative and Inductive. Collected Works of John Stuart Mill* (Volume VII). John Stuart Mill, ed. (Toronto: University of Toronto Press), 388–406.

Paine, Jack. Forthcoming. "Set-Theoretic Comparative Methods (STCM): Less Distinctive than Claimed." *Comparative Political Studies*.

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

Ragin, Charles. 2000. *Fuzzy-Set Social Science*. Berkeley: University of California Press.

Ragin, Charles C. 2005. "Core Versus Tangential Assumptions in Comparative Research." *Studies in Comparative International Development* 40: 33–38.

Ragin, Charles C. 2014. "Comment: Lucas and Szatrowski in Critical Perspective." *Sociological Methodology* 44: 80–94.

Ragin, Charles C. and Benoît Rihoux. 2004. "Replies to Commentators: Reassurances and Rebuttals." *Qualitative Methods* 2(2): 22–24.

Rihoux, Benoît. 2003. "Bridging the Gap Between the Qualitative and Quantitative Worlds? A Retrospective and Prospective View on Qualitative Comparative Analysis." *Field Methods* 15(4): 351–365.

Schneider, Carsten Q. and Claudius Wagemann. 2006. "Reducing Complexity in Qualitative Comparative Analysis (QCA): Remote and Proximate Factors and the Consolidation of Democracy." *European Journal of Political Research* 45(5):751–786.

Seawright, Jason. 2005. "Qualitative Comparative Analysis vis-a-vis Regression." *Studies in Comparative International Development* 40(1): 3–26.

Seawright, Jason. 2013. "Warrantable and Unwarrrantable Methods: The Case of QCA." Prepared for presentation at the Annual Meeting of the American Political Science Association, Chicago, August 2013.

Seawright, Jason. 2014. "Limited Diversity and the Unreliability of QCA." *Sociological Methodology* 44(1): 118–121.

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

Thiem, Alrik. 2014. "Mill's Methods, Induction and Case Sensitivity in Qualitative Comparative Analysis: A Comment on Hug (2013)." *Qualitative & Multi-Method Research* 12(2): 19–24.

Clarifying Misunderstandings, Moving Forward: Towards Standards and Tools for Set-Theoretic Methods

Ingo Rohlfing
University of Bremen, Germany
rohlfing@bigsss.uni-bremen.de

Carsten Q. Schneider
Central European University, Hungary
schneiderc@ceu.hu

Critical debates can push boundaries if they are open-ended in nature. In this spirit, we appreciate David Collier's efforts in putting together a set of articles about set theory and QCA in a symposium published in the previous issue of this newsletter. Many important issues about the principles and current practices of set-theoretic methods are raised and at least some contributors seem open to the possibility that set theory and QCA are worthy of being pursued and improved. We are grateful to the editor of the newsletter, Robert Adcock, for inviting us to respond to these arguments—an invitation we accept based on the expectation that the goal of everyone involved in this exchange is to strengthen set-theoretic methods rather than prematurely dismissing them based on what we find to be shaky arguments. We further believe that this can only be the start of a larger, open debate aimed at resolving misunderstandings and enhancing set-theoretic research.[1]

Since we cannot address every aspect of the contributions to the previous issue (henceforth only 'contributions'), we focus on what we perceive to be the most salient topics for the current assessment and future development of set-theoretic methods. We first outline points of agreement and briefly indicate what we see as the major points of disagreement. The disagreements are then discussed in more detail in the following sections. Our major response is that several contributions identify important topics and problems, but they take their point too far or even in the wrong direction. Some contrib-

[1] *Editor's Note*. This exchange will continue in the next issue of the QMMR newsletter, when David Collier will respond to Rohlfing and Schneider's response to the newsletter symposium he organized. The authors would like to thank Benoît Rihoux and Alrik Thiem for comments on an earlier version of this text.