



A qualitative assessment of QCA: method stretching in large-N studies and temporality

Victoria Finn^{1,2} 

Accepted: 8 November 2021
© The Author(s) 2022

Abstract

Qualitative Comparative Analysis (QCA) is a descriptive research method that can provide causal explanations for an outcome of interest. Despite extensive quantitative assessments of the method, my objective is to contribute to the scholarly discussion with insights constructed through a qualitative lens. Researchers using the QCA approach have less ability to incorporate and nuance information on set membership as the number of cases grows. While recognizing the suggested ways to overcome such challenges, I argue that since setting criteria for membership, calibrating, and categorizing are crucial QCA aspects that require in-depth knowledge, QCA is unfit for larger-N studies. Additionally, I also discuss that while the method is able to identify various parts of a causal configuration—the ‘what’—it falls short to shed light on the ‘how’ and ‘why,’ especially when temporality matters. Researchers can complement it with other methods, such as process tracing and case studies, to fill in these missing explanatory pieces or clarify contradictions—which begs the question of why they would also choose to use QCA.

Keywords Qualitative comparative analysis QCA · Qualitative methods · Case-oriented research · Causality · Temporality · Large-N studies

1 Introduction

Researchers are interested in explaining the—often complex, combinational, and conjunctural—causes of specific events (Ragin 2008; Schneider and Wagemann 2012). To do so, some scholars opt for the comparative method, which is a qualitative, set-theoretic, and case-oriented approach. They work with small populations, defining the boundaries of the study using macrosocial units, then choose cases and measure similarities and differences on theoretically relevant conditions.

Qualitative Comparative Analysis (QCA) was first envisioned by Charles C. Ragin in 1987, then further developed by Ragin in 2008 and with other scholars in a 2009 edited

✉ Victoria Finn
victoria.finn@eui.eu

¹ Max Weber Programme- Robert Schuman Centre for Advanced Studies, European University Institute, Fiesole, Italy

² Political Science Department, Universidad Diego Portales, Santiago, Chile

book. It is simultaneously qualitative and quantitative, taking advantage of the strengths associated with each tradition (Ragin 2008). QCA expounds that combining necessary and sufficient conditions in various cases, and then comparing them, can lead researchers to determine causal explanations for an outcome (Hug 2013: 253). The method is insensitive to how frequently the cause, or combinational causes, occurs for the outcome. This means fewer, but more in-depth, cases that demonstrate the phenomenon of interest, since one does not have to search for repetitive occurrences throughout a population. QCA allows researchers to explain complex complexity within configurations (e.g., equifinality) and to perceive the ‘nuances’ of necessary and sufficient causes (Aus 2009; Olsen 2014).

Despite the method’s original intention of analyzing only a small number of cases, later Ragin (2008: 7) extended QCA’s applicability also to large-N studies. When determining multiple and complex causality, Ragin (2008: 82) suggests that fuzzy sets allow researchers to reap the best from both quantitative and qualitative approaches. Whereas other scholars have focused on the quantitative shortcomings of QCA, this discussion focuses on its qualitative aspects. While I recognize the merits of QCA and its extensive use, the present article focuses on some of its limitations. I avoid responding to a ‘side’ (advocates versus opponents) but rather, in the spirit of participating in the scholarly debate about the method, the intent is to contribute insights constructed through a qualitative, more so than a quantitative, lens.

I consider the extension to large-N studies *method stretching* since as the number of cases of analysis grows, researchers using QCA have less ability to incorporate detailed information about each case. Larger-N studies complicate setting criteria for membership, calibrating, and categorizing, which in turn affects the conclusions—thus should not be advocated for or presented as feasible.¹ Using QCA for large-N studies makes poor use of a researcher’s in-depth case knowledge (Rutten 2020)—one of the main benefits of qualitative research. Goertz and Mahoney (2012: 1) start *A Tale of Two Cultures* by outlining quantitative and qualitative research as “loosely integrated traditions,” each with their “own values, beliefs, norms” and “distinctive research procedures and practices.” Inferences arise from cross-case analysis in quantitative research but from within-case analysis in qualitative studies, both which can be exploited in multimethod research (Goertz 2017; Goertz and Mahoney 2012).

While I review some of the ways to overcome this barrier while still using QCA (e.g., collaborating or including external experts), I nonetheless point to shortcomings of such workarounds, as well as possible heterogeneous conclusions and decreased rigor. Differing conclusions do not necessarily indicate they are incorrect, for example when researchers analyze different populations, periods, or contexts. I also discuss that while QCA can identify various parts of a causal configuration (the ‘what’), given that one of QCA’s main aims is to link observed patterns to existent theory, a useful part of the answer would also

¹ Ragin (2008: 208) has specifically addressed calibration for fuzzy sets, but he frames it as a strength: “Miscalibrations distort the results of set-theoretic assessments. The main principles guiding calibration are that (1) the target set must be carefully defined and labeled and (2) the fuzzy set scores must reflect external standards based on both substantive knowledge and the existing research literature. While some might consider the influence of calibration decisions ‘undue’ and portray this aspect of fuzzy-set analysis as a liability, in fact it is a strength. Because calibration is important, researchers must pay careful attention to the definition and construction of their fuzzy sets, and they are forced to concede that substantive knowledge is, in essence, a prerequisite for analysis.” Given this excerpt, I consider that “substantive knowledge” as a “prerequisite” strongly indicates that QCA should not be expanded to large-N studies.

explain *how* or *why* the causes contribute to the effect. From this standpoint, other methods such as process tracing and case studies can better offer such explanations.

The following Sect. 1.1 contains a review of QCA's strengths and variants, as well as some critiques of it, underlying how the present discussion can contribute to the dialogue. Section 1.2 outlines the main argument that both analytically and practically questions using QCA with large-N cases. The briefer Sect. 1.3 suggests that other methods may be more appropriate when temporality is key to answering the research question at hand, followed by the conclusion.

1.1 QCA: merits, types, and limits

As a broad introduction to its merits, QCA was originally a variation of the comparative method, is grounded in set theory, and is ideally suited for studying “explicit connections” between causes and effects (Ragin 2008: 23). Its Boolean algebra approach allows researchers to uncover combinations of causes that produce an outcome. QCA can be used to assess necessary and sufficient conditions as well as pinpoint multiple conjunctural causation patterns (Ragin 1987: 101). Since the first application published in a journal in 1984, QCA's popularity has grown—especially since 2004—most often used in political science, then sociology and anthropology, as well as in economics and management, and in multidisciplinary studies (Rihoux et al. 2013). Its use has continued to spread, although not without errors in effectively using the method (see Schneider and Wagemann [2010] for good practices and Rubinson and colleagues [2019] for common mistakes).

The method can be used to summarize data, develop new theoretical arguments, and check existent theories (Berg-Schlosser, De Meur, Rihoux, and Ragin 2009: 15). Classifying through codification can assist scholars in organizing a social phenomenon thus can also be beneficial for exploring and selecting cases, as well as answering descriptive research questions. Finally, the classification process can be considered transparent, as it allows researchers to control conditions (bounds) if they clearly explain concepts, attributes, and indicators.

There are three main types of QCA: crisp set, multi-value, and fuzzy set (csQCA, mvQCA, fsQCA, respectively) (see, e.g., Rihoux 2006; Schneider and Wagemann 2012; also Annex Fig. 1). The difference between the various forms lies in how to score (i.e., categorize) the concepts of interest. Scholars later developed another type of QCA, temporal QCA (TQCA), to attempt to overcome critiques that had been made regarding the method's inability to include temporality into its analysis (Caren and Panofsky 2005; Ragin and Strand 2008).

In its most basic and original form, csQCA, each case is classified as 0 (absence) or 1 (presence of the binary variable of interest) (Rihoux and De Meur 2009: 34–36). Using democracy as an example, a 0 may indicate non-democracy, whilst 1 is democracy. The second type, mvQCA, broadens the strict dichotomous nature. MvQCA extends csQCA to allow for more notation values; the threshold justifications should be based on theory or empirics (Cronqvist and Berg-Schlosser 2009: 70, 76). In other words, the researcher defines the extent to which each case is part of a subset. Continuing with the example of democracy, thinking along the lines of Collier and Mahon (1993), participatory democracy would have a notation value of 1, liberal democracy 2, and popular democracy 3.

The third type, fsQCA, is fuzzy-set scoring in which ‘fuzziness’ conveys the idea of conceptual boundaries that are not sharply defined (Schneider and Wagemann 2012: 27). It is fuzzy around the edges since the score ranges between 0 and 1. It seeks to determine

the degree of membership to a group, with 0 as absolute non-membership and 1 as full membership. In the democracy example, a case in fsQCA with a score of 1 would be a full-fledged democracy (following the researcher's precise definition of what that entails) whereas a case scoring 0.4 is a partial, even weak, democracy (Schneider and Wagemann 2012: 29). As compared to, for example, csQCA, Rihoux (2006) positions fsQCA as more appropriate for larger-N studies.

In addition to these three main types, TQCA tries to incorporate temporality, specifically by capturing the temporal nature of causal interactions (Caren and Panofsky 2005: 147). This type gives weight to the sequence—as in the temporal order—of case attributes that could be causally relevant for the outcome of interest; TQCA is bounded by theoretical restrictions to maintain a manageable number of possible configurations (Caren and Panofsky 2005: 148).

Quantitative scholars have put forth numerous critiques regarding problematic (particularly causal) inference, measurement errors, and fuzzy-set scoring which, they argue, doubt QCA's reliability and usefulness as a method (see, e.g., Bowers 2014; Braumoeller 2014; Collier 2014; Hug 2013; Krogslund and Michel 2014; Lieberman 2004; Lucas and Sztatrowski 2014; Munck 2016; Seawright 2005, 2014). For instance, some researchers state that QCA algorithms, which are its main analytical procedure, generate inconsistent results when applying the method repeatedly in simulations (Collier 2014; Krogslund and Michel 2014; Lucas and Sztatrowski 2014). QCA advocates have responded to such critiques (e.g., Berg-Schlosser et al. 2009; De Meur et al. 2009; Ragin and Strand 2008; Rihoux 2006; Rubinson et al. 2019; Schneider and Wagemann 2012).

However, in the qualitative literature, various issues within the QCA approach are presented as 'challenges' to overcome; the present article differs since it focuses on exploring a few of these qualitative shortcomings and suggests that the method should not be extended to large-N studies. To make these assessments, I combine the method's logic and cognitive science to provide a cognitive-epistemic and linguistic qualitative critique of QCA. Specifically, I build from literature based on cognitive science focused on categorization (Elkins 2014; Rosch 1975a, b, 1978; Zadeh 1965), linguistics (Lakoff 1975, 2014), logic (Munck 2016; Sartori 1970, 2014), and on temporality based on path dependency (Mahoney 2000; Pierson 2004). While the contributions support some of the quantitative critiques' takeaways, my parallel findings stem from a qualitative viewpoint—all the more reason that the discussion is of interest to readers of *Quality & Quantity*.

1.2 QCA: unfit for large-N studies

The main point is that when using Qualitative Comparative Analysis, sizes matters. In-depth knowledge intrinsically relates to smaller-N studies. Therefore, I argue that despite existent suggestions to use QCA in large-N studies, the method is only appropriate for small- and medium-N studies. While some types of cases are more scalable than others, as the number of cases in an analysis increases, the depth of a researcher's case knowledge is forfeited for breadth. In Rihoux and Ragin's (2009: 176) words, "as the number of cases grows, it becomes increasingly difficult to develop a sufficient knowledge of all individual cases." They clarify that "there must be sufficient 'case-based knowledge' before engaging in the further technical operations of QCA" but convey that researchers' main concern "should still be the original research question and the subsequent use of theory to guide case selection" (Rihoux and Ragin 2009: 24). While agreeing on the importance of the research question and use of theory, I also suggest that when the cases exceed a number

that eliminates the possibility to have ‘sufficient case-based knowledge,’ the researchers’ best option would be to select another method to appropriately answer their research question. While not arguing against the use of QCA completely, the proposed statement questions its use in large-N studies.

Fiss, Sharapov, and Cronqvist (2013: 191) recognize the conflict and its result, “in large-N QCA, it is difficult to maintain the kind of intimate familiarity with the cases that small-N QCA is usually based on. As a result, measurement errors in coding of cases are more likely.” For them, a step forward would be to combine large-N QCA with econometric analysis, convincingly outlining why the two methods, differing in their approach to social science research, can fruitfully be used in a hybrid method incorporating elements from both (Fiss et al. 2013: 194). I focus on the first issue they highlighted about researchers’ familiarity with their cases.

As one of the main strengths of qualitative research, a researcher’s in-depth case knowledge is diminished when using QCA (or other methods) beyond a small number of cases (Rutten 2020). This argument is two-tiered: as the number of cases grows, researchers have less capability to find and correctly incorporate in-depth information about each case, which lessens the ‘qualitativeness’ of a qualitative study. As Krogslund and Michel (2014: 25) point out, “the method relies heavily on the close knowledge of relatively few cases for making inferences.” Less specific knowledge, or empirical “intimacy” (Ragin 1994 cited in Rihoux and Ragin 2009: 24), results in risking inappropriate coding, which can lead to differing conclusions.

Two ways to overcome this would be to use thematic or country experts or to collaborate in a larger group of researchers (as suggested in Rihoux and Ragin [2009]). Analytically for any method, more collaborators could affect rigor and cohesiveness across cases; for QCA, relying on experts would mean accepting external assessment regarding the extent to which a case pertains to membership.² As Ragin repeatedly emphasizes, researchers return to the cases throughout the process of searching for causal configurations. The back-and-forth would be productive only when maintaining continued engagement with experts, to refine and recalibrate conditions and cases. Alternatively, the suggestion for collaboration is welcome, although finding researchers specialized in over 50 areas or countries would signify a large research project, rather than typical analyses. Such a strategy again seems appropriate for smaller-N studies and only for larger-N studies in specific projects.

Thereafter, this diminished in-depth case knowledge affects how the principal researchers are able to categorize each case correctly and uniformly. I of course recognize the richness of qualitative work that offers in-depth answers that contribute to a single part of the larger picture. Nonetheless, an aim is to be as scientific in methods as possible, and this means researchers should arrive at similar conclusions if they had asked similar research questions.

Ragin refers to quantitative studies as variable-oriented whereas qualitative research is case-oriented. Case-oriented implies that researchers know—meaning profoundly know—their cases. In-depth case knowledge is a defining principal benefit of qualitative studies. This critical point parallels what Gerring (2007: 10) states, “large-N cross-case analysis is always quantitative, since there are (by construction) too many cases to handle in a qualitative way.” Nonetheless, it seems to be unintendedly undermined, or at least restricted, in

² Analytical difficulties arise for much collaborative research and some projects overcome them—for example, by creating extensive databases across world regions. Such projects bring additional benefits to scholarly networks and individuals, such as collaborating with researchers in the Global South.

each of QCA's three main types (csQCA, fsQCA, and mvQCA), particularly while handling a larger number of cases (not synonymous with a large number of units of analysis).

Along with outlining perks and innovations of QCA as a research approach, Rihoux and Marx (2013) highlight parsimonious explanations as one of QCA's key benefits, while Aus (2009) positions QCA's strength in causal complexity over parsimony, for small-N studies. Connecting parsimony and causality, Baumgartner (2015: 840) highlights that "only maximally parsimonious solution formulas can represent causal structures" and as such, criticizes QCA researchers examining causal hypotheses who accept intermediate solution formulas (since parsimony is not maximized).³

In set-theory, one places an object (or case) into a category: e.g., Germany, India, and the United States are democracies. One can use Boolean algebra to classify each democracy using a score of one, or in the absence of democracy, a score of zero, then create a truth table listing all the cases in an ordered fashion (Ragin 1987: 86–88). In truth tables, each row is a case (or a possible logical combination of causes and outcome) and each column is a condition of interest (similar to independent variables), including the outcome (the dependent variable).

Researchers may alternatively choose to classify objects into partial categories, scoring between 0 and 1; or alternatively, for more dynamic classification, researchers can give partial membership based on degree and fuzzy-set theory, developed by Zadeh (1965). Lakoff (1975, 2014) extensively explains fuzzy logic, fuzzy concepts, and fuzzy-set scoring, as well as their relationships to natural language. Thereafter, Rosch (1975b) developed a prototype analysis allowing scholars to group cases according to characteristics and then calculate its membership based on the distance between the case and the prototype (also see Elkins [2014: 37]). This literature and line of thinking are key in understanding how to categorize using Ragin's fuzzy-set scoring.

The first issue is deciding which cases constitute an example of the phenomenon of interest (e.g., the example of democracy) while maintaining the concept's validity and without stretching the concept (Collier and Levitsky 1997; Collier and Mahon 1993; Goertz 2006; Sartori 1970). The second challenge is deciding to what extent each one is a democracy. There is no fixed answer for this question since it incorporates a matter of degree, or in other words, a degree of truth (Lakoff 1975; Rosch 1975b). This distorts the perceived clean image of how much truth a truth table may contain.

One of QCA's benefits is allowing the researcher to first set the boundaries and then place the object into (or, when working with fsQCA, partially into) the membership. Nevertheless, this is also a drawback of QCA since it can give false positives, or in other words, be the result of chance (for details, see Braumoeller [2015]). Since researchers can fall into confirmation bias in fuzzy sets (Krogslund, Choi, and Poertner 2015), they would interpret concepts differently, similar to how conceptual innovation (i.e., interpretation) has occurred with democracy.

The 'fuzziness' membership is scored from 0 to 1 (just as Zadeh [1965] proposed). But what fits, or does not fit, into membership of a concept? While utilizing fuzzy sets, the answer is concerning for scholars such as Sartori (2014: 15). Allowing researchers to set both the bounds and the membership within them makes QCA intrinsically inductive to use and leads to problematic inferences (Hug 2013). Yet in Ragin's edited volume, the authors

³ However, Baumgartner (2015) warns that if a QCA researcher must accept flawed simplifying assumptions to maximize parsimony, a better solution would be to use Coincidence Analysis as the optimization algorithm (instead of the Quine-McCluskey optimization used in csQCA and fsQCA).

of one chapter retort that researchers setting the thresholds is not a weakness, but rather allows for exploration to discover what occurs when the boundaries are slightly shifted; although they also state that in both csQCA and mvQCA, "... the results derived *might* depend on the thresholds selected, and therefore the thresholds should be selected with care" (Cronqvist and Berg-Schlusser 2009: 76, emphasis in original).

A certain degree of induction is suitable for qualitative research. Even though social scientists must remain open to inductive findings, qualitative research cannot rely solely on such discoveries since this could be a symptom of a poorly developed or interpreted theory. Ragin (1987: 42) states that his inductive approach can indeed handle multiple or conjectural causation (which, for instance, Mill's method of agreement and indirect method of difference cannot), and as a result, QCA is a more advanced technique than Mill's methods (Thiem 2014).

Moreover, fuzzy-set categorization is not a free-for-all. Ragin (1987) insists that membership must be calibrated: measuring devices are matched to known standards, based on theoretical concepts; calibration can be done directly or indirectly (see Ragin [2008], Chapter 5). Two decades later, it also included empirical justification, with warnings of transparently justifying the thresholds so that the study is understandable and replicable (Cronqvist and Berg-Schlusser 2009: 76). The benefit of fuzzy-set scoring is to avoid "black or white" dichotomies since it allows for, "more fine-grained assessment of set membership" (Rihoux and Marx 2013: 169). Sticking to the focus on larger-N studies, accurately choosing conditions and determining membership is a greater feat as the cases increase.

To replicate, other scholars would have to agree with the design and justifications of conditions and thresholds (reflecting previous conceptualizations and definitions), as well as interpretation, and check that the analysis had passed a benchmark test.⁴ As Rubinson and colleagues (2019) underline, "successful calibration requires one to carefully reflect upon the nature of one's measures and their meaning." If other researchers use or replicate the study, they must also ponder the in-depth work behind, and meaning of, calibrations, making it more difficult to undertake in a large-N study.

People should be able to categorize with minimum cognitive effort (Rosch 1978), yet fuzzy-set scoring requires a thinking exercise to determine the degree of membership (Elkins 2014). The cognitive effort associated with QCA truth tables intrinsically also requires subjectivism when one contemplates how to categorize each case; researchers interpret and apply the selected theory and their case-specific knowledge. Even since the beginning, Ragin (1987: 162) recognized that creating *useful* truth tables is the most intellectually demanding part of the method.

While qualitative case-oriented researchers dedicate time and effort to coding, the ensuing process of categorizing requires further cognitive effort since researchers need to amend (or recalibrate) categorizations as part of the process. While supporting membership categorization via necessary and sufficient criteria to define well-bounded concepts, Sartori (2014) rejects set theory as a dominant framework because it may bog researchers down in unproductive techniques (Collier 2014: 4). Other researchers would have to agree with the series of conditions and set relations included in the study before being able to verify its results.

⁴ Marx and Dusa (2011) developed benchmark tables to "assess the chances of accepting a model for further analysis" which determines if researchers using a QCA approach would be able to "distinguish an analysis on the basis of random data versus real data" (as reviewed in Rihoux et al. [2013: 180]).

Considering categorization, I turn to Goertz's (2006: 6) conceptual construction, positioning social science concepts as both multidimensional and multilevel. When working on a concept's basic level, a scholar must pay attention to three issues: the negative pole, the continuity that exists (or not) between the poles, and the content of the continuum between the two poles, or the "gray zone" (Goertz 2006: 30). Conceptualizing in csQCA falls short: working with 0 and 1, it only allows for including the positive pole (the concept itself) and the negative pole. The negative pole is only completely addressed when scholars explicitly explain it, including defining the substantive content of 0.

FsQCA indeed allows for exploring the gray area since the scholar classifies based on degrees of membership to the positive pole. Continuing using Goertz's (2006: 41) words, fuzzy logic is fitting for continuous variables "since it is an infinite-valued logic." Nevertheless, the location does not help to understand how the case relates to the social world beyond the truth table, unless the researcher also explains the contents of the negative pole as well as the gray zone. For instance, if the concept is democracy and case (A) has a fsQCA score of 0.8, others understand that it is quite close to being a democracy. Yet, this information alone expresses little until more is known about the meaning of 0, which is the negative pole: is it autocracy, dictatorship, authoritarianism? Using asymmetrical calibration, it should be nondemocracy, which should be accompanied by a substantive explanation of its meaning (Rubinson et al. 2019). From there, others must be able to grasp the meaning of the values between 0 and 1 for each concept. In Goertz's (2006: 30) lingo, this would require explicitly explaining the "substantive content of the continuum between the two poles."

These points lay the groundwork to see that the rigorous mental exercise of scoring in QCA truth tables requires vast in-depth knowledge of the phenomenon, the cases, and the theory before a researcher can accurately justify its categorization either to membership (0 or 1) or to a degree of membership (somewhere between 0 and 1). It requires both cognitive effort and interpretation. The exercise then also results in epistemic danger and volatile, less credible conclusions. Inconsistency among studies fails to add to the social science body of knowledge, which reduces QCA's usefulness as a method as the number of cases grows.

Given epistemic and cognitive limitations, *in-depth knowledge is more unlikely as the sample size increases*. Initially, Ragin recognizes that case-oriented studies work best with about 2–4 positive cases, plus the same number of negative cases, and states that QCA is more difficult to use as the number of cases increases; specifically, that the approach is "incapacitated by a large number of cases" (Ragin 1987: 49, 69). Despite the method's original intention, Ragin (2008: 7) later extends it by saying that, "the set-theoretic methods I had developed for small-N and medium-N research could be productively extended to large-N." This is reiterated again by Ragin and co-authors (Berg-Schlusser et al. 2009: 17) and by QCA supporters, stating that analyzing a mid-sized number of cases does not violate any of the method's assumptions so it can also be used for analyzing large-N data (Schneider and Wagemann 2012: 13). Such expansions overstate the method's capacity and should not be advocated for as feasible.

Continuing this thinking, QCA with a larger number of cases "does not sacrifice explanatory richness" yet only applies to certain clusters and contexts (Rihoux 2006: 698). To be more specific regarding the numbers and approaches, Rihoux (2006: 686–687) classifies small-N situations as those with less than 30–40 cases, which emphasize case-based knowledge, and that best fit dichotomous QCA; medium-N situations are 40–50 cases and work best with mvQCA; and finally, fuzzy sets are best used in large-N situations, typically about 50–80 cases in practice (but some QCA studies have used over 100 cases) (see

Annex Fig. 1). Caren and Panofsky (2005: 151) state that successful studies have used between 18 and 50 cases.

Greckhamer, Misangyi, and Fiss (2013) recognize that increasing the number of cases in QCA changes the assumptions, objectives, and approach, suggesting that “two QCAs” exist: one for small-N and one for large-N (which they consider over 50 cases). Regarding the richness of information that the researcher has on each case, even Rihoux’s (2006: 686) consideration of small (30–40) demonstrates a large discrepancy from Ragin’s original 4–8 cases. This shows how quickly the ‘in-depth’ case knowledge would be degraded and how the cognitive burden of the classification drastically increases when the set is quadrupled or multiplied even further. This is along the lines of what I call *method stretching*, or applying the method beyond its useful capacity. One aim of this analysis is to urge advocates and scholars not to fall into method stretching with QCA.

With more generalized and less contextual knowledge, I argue that one cannot correctly—or at least with low risk of making incorrect claims—interpret concepts and theories to properly set membership thresholds, nor appropriately place cases in a fuzzy set. I have already addressed issues surrounding collaborative efforts, which affect many research projects, not just those applying QCA.

To reiterate, unjust or non-uniform categorization can create fragile categorical definitions, even those rooted in theory. The process is prone to subjectivity since scholars could justify a variety of categories, as long as they could provide a plausible interpretation of theory. Returning to Fiss and colleagues (2013: 191), “contradictory observations in large-N QCA might then at times be accepted as measurement error, whereas in small-N QCA, they will frequently trigger a re-examination of the cases selected and whether all relevant causal conditions have been included.” I interpret that producing heterogeneous conclusions undermine part of the method’s usefulness of building knowledge and consensus. Rather than considering such contradictions as measurement error, a researcher would want to further examine cases more in-depth.

When various contradictory cases appear, Greckhamer and colleagues (2013) suggest an in-depth analysis of a randomly selected sample of contradictory cases. But the main point stands that if researchers knew the phenomenon and cases, then it begs the question of why they would opt for QCA at the start rather than, for example, select one or more case studies. As Gerring (2007: 12, 89–90) points out, researchers engage in a cross-case approach to select case studies then choose an appropriate type, based on the research question and if they aim to generate or test hypotheses.⁵ Such methods link case characteristics and outcomes to theory, to question or refine existent theory.

1.3 When temporality matters

Despite suggestions of incorporating temporality into QCA, when temporality is key to understanding the question at hand, I suggest that other methods may be more appropriate. This is not the first time researchers have pointed out this weakness; it has been recognized in Ragin’s 2009 edited volume (De Meur et al. 2009: 161–163) and some scholars have tried to overcome it. At least two solutions have been put forth: first, combining QCA with other techniques involving temporality (Boswell and Brown 1999; Griffin 1992), such as

⁵ Gerring (2007: 88–90) describes nine cases study types: typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different.

time series (Hino 2009), or second, directly including temporality in QCA, as Caren and Panofsky (2005) proposed.

Countering the first suggestion, I would ask if researchers are using another method, why would they additionally use QCA? In a chapter co-authored by Ragin, the authors state that, “QCA can lay the groundwork and be extended to even more demanding types of analyses—for example, taking into account the temporal dimension and the various ‘paths,’ ‘critical junctures,’ and overall dynamics...” (Berg-Schlusser et al. 2009: 7). Here temporality is linked to the idea of path dependency, which is a dynamic process involving positive feedback, generating multiple possible outcomes depending on the sequence in which events unfold (Pierson 2004: 20). This is not to say that path dependence is the only type of temporality, nor that all processes are path dependent or affected by positive (or negative) feedback. Temporality means it is not just *what* happens but *when* it happens and *for how long* it happens. What happens first (and why) greatly matters for what happens next since previous happenings affect the following occurrences. A researcher with in-depth case knowledge understands these details within and between the selected cases. QCA fails to account for temporality, so researchers using the approach are missing an important part of the story. If using QCA to lay the groundwork, researchers could simply consider path dependence and a method such as process tracing (see e.g., Bennett and Checkel 2014; Mahoney 2012). As Goertz (2017: 49) explains, “the central purpose of process tracing is to find, verify, or disconfirm hypotheses about causal mechanisms.”

The second suggestion of how to include temporality into QCA comes from Caren and Panofsky (2005), calling it temporal QCA (TQCA). This addition deals with a type of temporality based on trajectory—meaning timing in a sequential order. Basing temporality on trajectory differs from path dependence. Path dependence involves positive feedback that generates multiple possible outcomes depending on the order in which events unfold (Pierson 2004: 20). Although path dependence more realistically reflects the social world, TQCA is limited to only simple cases due to restricting the number of possible configurations (Caren and Panofsky 2005: 163). Analyzing social phenomena involve more than the allowable sequences possible with TQCA. These restrictions mean that temporality has not been properly incorporated.

With added clarifications, Ragin and Strand (2008: 440) recommend TQCA while using “simple temporal sequences.” In TQCA, temporal order is included to capture the conditions that are “potentially causally relevant,” can be used with crisp and fuzzy data sets, is grounded in set theory, uses truth tables, and makes necessity and sufficiency statements (Schneider and Wagemann 2012: 16). Yet, due to QCA’s analytical use of truth tables, TQCA must be limited to combinations of only *two* temporal factors at a time to be considered one sequence and can empirically accommodate up to only *four* sequences (Ragin and Strand 2008: 439; Schneider and Wagemann 2012: 270). Two temporal factors are, for instance, having authoritarianism then democracy (in that order, which creates one sequence). But a researcher studying current democracies knows that having had a democracy, authoritarianism, then re-democratization will look different from a case that had nondemocracy, authoritarianism, then democracy. Hence allowing the researcher combinations of only two temporal factors are insufficient for adding temporality into QCA. Using a sequence of two could miss the chance of recognizing causal explanations since sequences of causally connected events occur under certain conditions and in a repeated manner (see Mayntz [2004]).

In limiting temporal factors to two in each sequence, TQCA further is limited to only four sequences. This seems overly restrictive since it forces the researcher to choose the most critical ones, while overlooking the others. Precisely to maintain a small number

of variables as Ragin intended, Caren and Panofsky (2005: 163) incorporated temporality based on trajectory, not path dependence. The difference is that trajectory is a path, but path dependence contains a particular event—for instance, a critical juncture—that can ‘derail’ the sequence into an alternative one (Caren and Panofsky 2005: 163; Gerring 2007; Mahoney 2000). Since “TQCA is only suitable for very simple instances of path dependency” (Caren and Panofsky 2005: 163), it falls short of capturing Pierson’s (2004) path dependence with self-reinforcing, positive feedback loops and the Polya urn process. Thus, when temporality matters, researchers should look to other methods.

To better understand the importance of temporality in QCA, I turn to a discussion on logic since it comprises a critical part of context-dependent categorization, directly affecting researchers’ causal conclusions. Set-theoretic comparative methods “reduce causation to a logical relation and erroneously posit that causal hypotheses can be formalized as a relation of material implication” (Munck 2016: 775). One cannot simply infer causal conclusions from associational relations (Paine 2016: 706). Logical relations are not synonymous to causal relations since the latter requires a change in X resulting in a change in Y, so causation should not be reduced to logical relations (Munck 2016: 777). Recoding a case in QCA can then produce different ‘causal’ conclusions (Goldthorpe 1997).

The logical semantics behind the degree of membership are also important. Lakoff (1975) uses terms such as technically, strictly speaking, loosely speaking, and regular to scale an item to a group; note that these are *non-linear*. This is unlike fuzzy logic and fuzzy set-scoring, which place the item on a line from 0 to 1, where it is possible to measure the numerical points between those values. The differences—or the distance—between Lakoff’s terms are unmeasurable: “strictly speaking” may be closer to “technically” than to “loosely speaking,” or vice versa. Thresholds are also used, but categorization occurs through more natural language: “strictly speaking” means that each criterion is above certain thresholds to be included in the membership and the boundaries are context dependent (Lakoff 2014: 10). To connect these ideas, context, history, and past occurrences in that specific context define the thresholds. To effectively categorize in QCA, researchers make measurement decisions based on their in-depth case knowledge. Where a case is placed depends on temporality within cases.

Alongside the temporality discussion, despite its ability to *explore* multicausality, QCA also falls short when trying to *explain* outcomes. Explanation includes the how and why those conditions in a certain context contribute to the outcome (Falleti and Lynch 2009). A causal mechanism consists of recurrent processes that connect initial conditions (causes) and that are composed of causally linked events that form sequences (Mayntz 2004: 239–243); moreover, context is not necessarily part of the causal relationship (Denk and Lehtinen 2014). Mechanism-based explanations have begun to draw more attention from scholars throughout the social sciences (Hedström and Ylikoski 2010: 49). Of course, not all research questions are causal; but when they are, causal mechanisms explain how various elements connect different causes to produce the outcome of interest (Y). In these situations, multimethod research (taking cross-case and within-case causal inference as complementary) “means a commitment to the causal mechanism approach to social and political research, which itself means a commitment to case studies as the methodology for exploring causal mechanisms” (Goertz 2017: 5, 29).

QCA can identify the different conditions (X_n) that compose a causal configuration, which determines the occurrence of an outcome (Y), but it does not account for how nor why each condition impacts the outcome of interest. So, QCA on its own is unable to understand the exact effect X has on Y, further complicated by not being able to prioritize the sequence or duration of events (links within the causal chain) has on an outcome

of interest. Following Goertz (2017), multimethod research inherently combines methodologies and as such, researchers can combine cross-case causal inference methods (QCA being one choice) with case studies to pinpoint and analyze causal mechanisms.

2 Conclusions

In the spirit of advancing research and supporting qualitative methods, I have argued against advocating for QCA in large-N studies because more cases reduce the researchers' own in-depth knowledge, a key benefit of qualitative, or *case-oriented*, research. In short, researchers should avoid method stretching by not using QCA in large-N studies. During analysis, researchers' bound setting and membership scoring directly affect the causal conclusions reached. Setting bounds in QCA creates a risk: in practice, if different researchers possess the same information and research question, they will produce different bounds—even using the same theories—resulting in varying memberships and then distinct causal explanations. Fuzzy-set membership can be estimated improperly, or differently: for instance, membership can be set at the mean rather than the minimum, the true threshold of the category may be off, which results in incorrect conclusions (Braumoeller 2014: 46). The process of fuzzy-set scoring can result in measurement error, which must not be ignored, bringing the researcher to misleading inferences (Hug 2013: 252). Using csQCA or mvQCA still requires determining bounds and inductive results. The scorings do not allow for temporality within and between cases. QCA (even TQCA) fails to adequately include temporality, weakening the 'qualitativeness' of QCA and it overlooks pieces of the case's specific context. Social scientists' conclusions must aim to build from existing information and generate new scientific knowledge. Varying and incomplete answers to a research question fail to reach this objective.

In addition to focusing on its limits from a qualitative perspective, I have also outlined many merits and uses of QCA. Scholars can use QCA productively for exploratory purposes for combinational causes of outcomes of interest (Berg-Schlosser et al. 2009: 17), descriptive analyses, or to test deterministic hypotheses (Hug 2013: 252). It can also provide valuable assistance when synthesizing results, particularly when a researcher is attempting to identify the causal configuration that determines the outcome of interest. Despite advocates' recommendations to pair QCA with another method, QCA applications from 1984 to 2011, Rihoux and colleagues (2013: 181) found 61.3% used QCA as a standalone tool. Since then, advances have been made in multimethod research, especially combining QCA with other methods for specific purposes. For example, in contextual analyses, Denk and Lehtinen (2014) outline how to combine csQCA and fsQCA with comparative multilevel analysis. Fiss and colleagues (2013) suggest steps toward creating hybrid methods combining aspects from the QCA approach and econometric methods.

As also discussed, QCA is able to find the 'what' (i.e., what combination of conditions generates the outcome of interest) and explore complex causality, but is unable to explain how and why those causes, within that particular combination, in that order, under certain circumstances, determine the occurrence of Y. Fear not: alternatives exist, for instance, MIMIC Modeling (Multiple-Indicators Multiple Causes), similarity-based measures, latent-class analysis (Elkins 2014), family resemblance (Rosch and Mervis

1975; Wittgenstein 1953), or vertical and horizontal dimensional categorization and classical versus radial subtyping (Collier and Levitsky 1997; Collier and Mahon 1993; Rosch 1978). For parsimony in causal hypotheses, Baumgartner (2015) suggests Coincidence Analysis. To find or explore causal mechanisms, another option is process tracing (e.g., Bennett 2008; Bennett and Checkel 2014; Mahoney 2012), which successfully navigates cases' context specificity and can address historical sequences or outcomes, as well as handle complex and conjunctural explanations of outcomes. To find evidence of causal mechanisms, Goertz (2017, Chapter 7) reviews how large-N qualitative testing has emerged—used for up to around 50 cases. Over this amount, as underlined in the present discussion, QCA is unfit for larger-N studies since setting criteria for membership, calibrating, and categorizing are crucial aspects requiring in-depth case knowledge.

Annex

See Fig. 1.

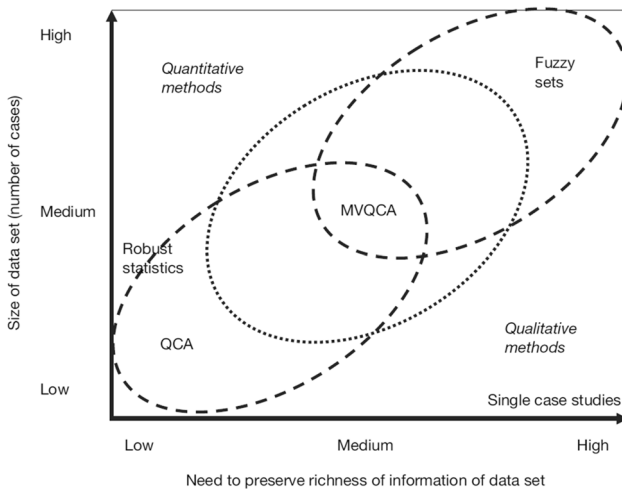


Fig. 1 Choosing a Method: Rihoux’s description between in-depth case knowledge and number of cases. *Source* Rihoux (2006), Fig. 1 Best Use of QCA, MVQCA and Fuzzy Sets. Here small-N studies are considered less than about 30–40 cases, medium-N are 40–50 cases, and large-N in practice have been between 50 and 80 but have included over 100 cases (Rihoux 2006: 686–687, 698). Rihoux’s “richness of information” is what I refer to as in-depth case knowledge.

Acknowledgements Thank you to Fernando Rosenblatt for the initial motivation and first review of this paper, to Gary Goertz and Luicy Pedroza for commenting on very early drafts, which greatly improved and balanced the arguments, and to Seba Koch for many fuzzy-set conversations. Many thanks also to the numerous anonymous reviewers for their patience and constructive rounds of comments.

Funding Financial support from COES (Centro de Estudios de Conflicto y Cohesión Social), CONICYT/FONDAP/15130009.

Availability of data and material Not applicable.

Code availability Not applicable.

Declaration

Conflict of interest The author reports no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aus, J.P.: Conjunctural causation in comparative case-oriented research. *Qual. Quant.* **43**, 173–183 (2009)
- Baumgartner, M.: Parsimony and causality. *Qual. Quant.* **49**, 839–856 (2015)
- Bennett, A.: Process tracing: a Bayesian perspective. In: Box-Steffensmeier, J., Brady, H.E., Collier, D. (eds.) *The Oxford Handbook of Political Methodology*, pp. 702–721. Oxford University Press, Oxford (2008)
- Bennett, A., Checkel, J.: *Process Tracing: from Metaphor to Analytic Tool*. Cambridge University Press, Cambridge (2014)
- Berg-Schlosser, D., De Meur, G., Rihoux, B., Ragin, C.C.: Qualitative comparative analysis (QCA) as an approach. In: Rihoux, B., Ragin, C.C. (eds.) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Applied Social Research Methods. Sage Publications, Thousand Oaks (2009)
- Boswell, T., Brown, C.: The scope of general theory: methods for linking deductive and inductive comparative history. *Sociol. Methods Res.* **28**(2), 154–185 (1999)
- Bowers, J.: Comment: Method games—a proposal for assessing and learning about methods. *Sociol. Methodol.* **44**(1), 112–117 (2014)
- Braumoeller, B.F.: Analyzing interactions: four alternative models. *Qual. Multi-Method Res* **12**(1), 41–45 (2014)
- Braumoeller, B.F.: Guarding against false positives in qualitative comparative analysis. *Polit. Anal.* **23**(4), 471–487 (2015)
- Caren, N., Panofsky, A.: TQCA. A technique for adding temporality to qualitative comparative analysis. *Sociol. Methods Res.* **34**(2), 147–172 (2005)
- Collier, D.: Problematic tools: introduction to symposium on set theory in social science. *Qual. Multi-Method Res.* **12**(1), 2–8 (2014)
- Collier, D., Levitsky, S.: Democracy with adjectives. *W. Pol.* **49**(3), 430–451 (1997)
- Collier, D., Mahon, J.E. Jr.: Conceptual 'stretching' revisited: adapting categories in comparative analysis. *Am. Polit. Sci. Rev.* **87**(4), 845–855 (1993)
- Cronqvist, L., Berg-Schlosser, D.: Multi-value QCA (mvQCA). In: Rihoux, B., Ragin, C.C. (eds.) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Applied Social Research Methods, pp. 69–86. Sage Publications, Thousand Oaks (2009)

- De Meur, G., Rihoux, B., Yamasaki, S.: Addressing the critiques of QCA. In: Rihoux, B., Ragin, C.C. (eds.) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. Applied Social Research Methods, pp. 147–166. Sage Publications, Thousand Oaks (2009)
- Denk, T., Lehtinen, S.: Contextual analyses with QCA-methods. *Qual. Quant.* **48**, 3475–3487 (2014)
- Elkins, Z.: Measuring Partial membership in categories: alternative tools. *Qual. Multi-Method Res* **12**(1), 33–40 (2014)
- Falletti, T.G., Lynch, J.F.: Context and causal mechanisms in political analysis. *Comp. Polit. Stud.* **42**(9), 1143–1166 (2009)
- Fiss, P.C., Sharapov, D., Cronqvist, L.: Opposites attract? Opportunities and challenges for integrating large-N QCA and econometric analysis. *Polit. Res. Q.* 191–198 (2013)
- Gerring, J.: *Case Study Research: principles and practices*. Cambridge University Press, Cambridge (2007)
- Goertz, G.: *Social science concepts: a user's guide*. Princeton University Press, Princeton (2006)
- Goertz, G.: *Multimethod research, causal mechanisms, and case studies: an integrated approach*. Princeton University Press, Princeton (2017)
- Goertz, G., Mahoney, J.: *A Tale of two cultures: qualitative and quantitative research in the social sciences*. Princeton University Press, Princeton (2012)
- Goldthorpe, J.H.: Current issues in comparative macrosociology: a debate on methodological issues. *Comp. Soc. Res.* **16**, 1–26 (1997)
- Greckhamer, T., Misangyi, V.F., Fiss, P.C.: The Two QCAs: From a small-N to a large-N set theoretic approach. In: Fiss, P.C., Cambré, B., Marx, A. (eds.) *Configurational Theory and Methods in Organizational Research (Research in the Sociology of Organizations, vol. 38, pp. 49–75*. Emerald Group Publishing Limited, Bingley (2013)
- Griffin, L.: Temporality, events, and explanation in historical sociology. *Sociol. Methods Res.* **20**(4), 403–427 (1992)
- Hedström, P., Ylikoski, P.: Causal mechanisms in the social sciences. *Annu. Rev. Sociol.* **36**, 49–67 (2010)
- Hino, A.: Time-Series, Q.C.A.: Studying temporal change through Boolean analysis. *Sociol. Theory Methods* **24**(2), 247–265 (2009)
- Hug, S.: Qualitative comparative analysis: how inductive use and measurement error lead to problematic inference. *Polit. Anal.* **21**(2), 252–265 (2013)
- Krogslund, C., Choi, D.D., Poertner, M.: Fuzzy sets on shaky ground: parameter sensitivity and confirmation bias in fsQCA. *Polit. Anal.* **23**(1), 21–41 (2015)
- Krogslund, C., Michel, K.: Where do we go from here? A larger-N, fewer variables problem? The counterintuitive sensitivity of QCA. *Qual. Multi-Method Res* **12**(1), 25–32 (2014)
- Lakoff, G.: Hedges: A study in meaning criteria and the logic of fuzzy concepts. In: Hockney, D., Harper, W., Freed, B. (eds.) *Contemporary Research in Philosophical Logic and Linguistic Semantics*. Springer, Dordrecht (1975)
- Lakoff, G.: Set theory and fuzzy sets: their relationship to natural language. Interview. *Qual. Multi-Method Res* **12**(1), 9–13 (2014)
- Lieberson, S.: Comments on the use and utility of QCA. *Qual. Methods Newsl. Am. Polit. Sci. Assoc. Org. Sect. Qual. Methods* **2**(2), 13–14 (2004)
- Lucas, S.R., Sztatowski, A.: Qualitative comparative analysis in critical perspective. *Sociol. Methodol.* **44**(1), 1–79 (2014)
- Mahoney, J.: Path dependency in historical sociology. *Theory Soc.* **29**(4), 507–548 (2000)
- Mahoney, J.: The logic of process tracing tests in the social sciences. *Sociol. Methods Res.* **41**(4), 570–597 (2012)
- Marx, A., Dusa, A.: Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification. *Methodol. Innov. Online.* **6**(2), 103–148 (2011)
- Mayntz, R.: Mechanisms in the analysis of social macro-phenomena. *Philos. Soc. Sci.* **34**(2), 237–259 (2004)
- Munck, G.L.: Assessing set-theoretic comparative methods: a tool for qualitative comparativists? *Comp. Polit. Stud.* **49**(6), 775–780 (2016)
- Olsen, W.: Comment: The usefulness of QCA under realist assumptions. *Sociol. Methodol.* **44**(1), 101–107 (2014)
- Paine, J.: Set-theoretic comparative methods: less distinctive than claimed. *Comp. Polit. Stud.* **49**(6), 703–741 (2016)
- Pierson, P.: *Politics in Time. History, Institutions, and Social Analysis*. Princeton University Press, Princeton (2004)
- Ragin, C.C.: *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, Berkeley (1987)

- Ragin, C.C.: *Constructing Social Research. The unity and diversity of method*. Pine Forge Press, Newbury Park (1994)
- Ragin, C.C.: *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press, Chicago (2008)
- Ragin, C.C., Strand, S.I.: Using qualitative comparative analysis to study causal order: comment on Caren and Panofsky (2005). *Sociol. Methods Res.* **36**(4), 431–441 (2008)
- Rihoux, B.: Qualitative comparative analysis (QCA) and related systematic comparative methods recent advances and remaining challenges for social science research. *Int. Sociol.* **21**(5), 679–706 (2006)
- Rihoux, B., Álamos-Concha, P., Bol, D., Marx, A., Rezsöházy, I.: From Niche to mainstream method? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011. *Polit. Res. Q.* 175–184 (2013)
- Rihoux, B., De Meur, G.: Crisp-set qualitative comparative analysis (csQCA). In: Rihoux, B., Ragin, C.C. (eds.) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. *Applied Social Research Methods*, pp. 33–68. Sage Publications, Thousand Oaks (2009)
- Rihoux, B., Marx, A.: Qualitative comparative analysis at 25: state of play and agenda. *Polit. Res. Q.* 167–171 (2013)
- Rosch, E.: Cognitive reference points. *Cogn. Psychol.* **7**(4), 532–547 (1975a)
- Rosch, E.: Cognitive representations of semantic categories. *J. Exp. Psychol.* **104**(3), 192–233 (1975b)
- Rosch, E.: Principles of categorization. In: Rosch, E., Lloyd, B.B. (eds.) *Cognition and Categorization*. Lawrence Erlbaum, Hillsdale (1978)
- Rosch, E., Mervis, C.B.: Family resemblances: studies in the internal structure of categories. *Cogn. Psychol.* **7**(4), 573–605 (1975)
- Rubinson, C., Gerrits, L., Rutten, R., Greckhamer, T.: Avoiding common errors in QCA: a short guide for new practitioners. *Sociol.* **9**, 397–418 (2019)
- Rutten, R.: Applying and assessing large-N QCA: causality and robustness from a critical realist perspective. *Sociol. Methods Res.* (2020)
- Sartori, G.: Concept misformation in comparative politics. *Am. Polit. Sci. Rev.* **64**(4), 1033–1053 (1970)
- Sartori, G.: Logic and set theory: a note of dissent. *Qual. Multi-Method Res* **12**(1), 14 (2014)
- Schneider, C.Q., Wagemann, C.: Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. *Comp. Sociol.* **9**(3), 397–418 (2010)
- Schneider, C.Q., Wagemann, C.: *Set-Theoretic Methods for the Social Sciences: a Guide to Qualitative Comparative Analysis*. Cambridge University Press, New York (2012 [2007])
- Seawright, J.: Assumptions, causal inferences, and the goals of QCA. *Stud. Comp. Int. Dev.* **40**(1), 39–42 (2005)
- Seawright, J.: Comment: limited diversity and the unreliability of QCA. *Sociol. Methodol.* **44**(1), 118–121 (2014)
- Thiem, A.: Mill's methods, induction and case sensitivity in qualitative comparative analysis: a comment on Hug (2013). *Qual. Multi-Method Res.* **12**(2), 19–24 (2014)
- Wittgenstein, L.: *Philosophical Investigations*. Blackwell Publishing, Oxford (2001 [1953])
- Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**(3), 338–353 (1965)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.