

# **Taming Limited Diversity and Causal Imbalance: A Two-Step Approach for Multi-Value Qualitative Comparative Analysis (mvQCA)**

**Tom Mannewitz, Technical University Chemnitz**

**Paper presented at the QCA Expert Seminar 26 April 2013 in Budapest**

## **Introduction**

Since its development by Charles C. Ragin in 1987 Qualitative Comparative Analysis (QCA) has been enjoying growing popularity and innumerable enhancements; first and foremost fuzzy set QCA (fsQCA). Even though it is not accepted by all users as an equipollent approach to csQCA or fsQCA yet (Vink and van Vliet 2007) multi-value QCA (mvQCA) – primarily advanced by Lasse Cronqvist (2007) – is another extension of Ragin’s first proposals. By allowing for the use of multinomial (mvQCA) and fuzzy conditions (fsQCA) both versions have overcome the dichotomy characteristic for csQCA. However, there is still a number of reasons for which the qualitative comparative method is exposed to criticism.

Representatives of one of the most strenuously defended position argue that QCA’s major shortcoming is: The causal statements derived from the analysis may rest on simplifying assumptions about logical remainders and thus be devoid of solid empirical foundation. This holds true even more for analyses that rest on fuzzy and multinomial conditions (Schneider and Wagemann 2007:265). Logical remainders are combinations of conditions that have not occurred but are logically possible. This problem is known as “limited diversity” (Ragin 1987:104-113). In order to handle it Carsten Schneider and Claudius Wagemann (2003; 2006; 2007:256-262) have designed a two-step approach that would allow for an immense reduction of logical remainders by analysing sufficient conditions in two separate steps. The approach is in part inspired by Herbert Kitschelt’s ideas about deep and shallow explanations of post-communist regime change (Kitschelt 1999; Kitschelt 2003) but enlarges it with own ideas and thereby makes it fruitful for a use within QCA. What speaks for the approach – besides its methodological benefit – is that it enables causal mechanisms by taking into account both causal deep and shallow conditions. It thus strikes a balance between various causal depths of explanatory factors. The benefits of the two-step approach are – however – still reserved for fsQCA (Schneider and Wagemann 2007:262) even though applications of a rudimental mvQCA two-step approach exist (Mannewitz 2011; Sager and Andereggen 2012). The present contribution aims at elaborating the module for mvQCA in a systematic manner.

The paper is divided into three sections: The first part introduces the problem of limited diversity and several counter measures. The second one illustrates Schneider and Wagemann’s approach – its technical procedure, its function logic, and the reasons that speak for its application. An introduction to QCA cannot be achieved here. The length of this section is due to the fact that it will lay the foundation of the second part. It illustrates core aspects that speak in favour of the development of an mvQCA two-step approach. Schneider and Wagemann’s module is most helpful both in reducing the potential number of logical remainders *and* enlightening the study of social phenomena as it arranges analysing steps into

separate phases: the consideration of remote conditions first, the inclusion of proximate conditions afterwards.

In the third part I am going to derive an mvQCA two-step approach from the existing fsQCA module presented by Schneider and Wagemann. This section is structured similarly to the second section – it thus includes the presentation of conduction, the function logic, and the benefits of an mvQCA two-step approach. For better understanding several aspects deserve a comparison to the fsQCA two-step approach.

### **The problem of limited diversity and counter measures**

QCA allows for the investigation of sufficient and necessary configurations of conditions in middle-sized case sets with recourse to set theory. In the first version – nowadays known as crisp set QCA – the conditions can only take dichotomous values: present (1) or absent (0). From this starting point a truth table – based on the empirical information of the present data set – is generated in which each row represents one unique combination of differently shaped conditions – not necessarily cases, since different cases might display the same combination of conditions. By using a bottom-up process or the Quine-McClusky-Algorithm the information of this table are minimised (Schneider and Wagemann 2007:49-73; Ragin 1987:93-102; Rihoux and Ragin 2009:56-64) into a solution formula that displays sufficient and necessary conditions respectively and omits conditions not relevant for the outcome.

A major downside of QCA is the problem of simplifying assumptions on logical remainders (Ragin 1987:104-112; Schneider and Wagemann 2007:101-115). They appear whenever logical remainders, i. e. logically possible combinations of conditions without empirical evidence, are included into the minimisation process (Schneider and Wagemann 2007:101). Three reasons can be named for logical remainders (Schneider and Wagemann 2012: 154-157):

- Impossible remainders: There are logically possible configurations, but there is no empirical evidence for them since they are empirically impossible – e. g. pregnant men or lactating birds.
- Clustered remainders: There are logically possible configurations, but there is no empirical evidence for them since they have not occurred yet – e. g. female US presidents.
- Arithmetic remainders: Certain configurations are represented by cases. However the case selection does not contain them.

Depending on what outcomes such combinations would evoke the results of a certain QCA may vary. There are different strategies that try to cope with the problem:

- 1) Ceding the creation of the most parsimonious solution to the software's simulation processes – this may lead to the inclusion of numerous simplifying assumptions on logical remainders.
- 2) Eliminating logical remainders from the minimisation process – by doing so the solution formula will exclusively rest on information the truth table contains.

- 3) Coding certain not occurring configurations according to own theoretical considerations on various scenarios. This proceeding requires some good reasons to code a configuration in this (truth value: 0) or that (truth value: 1) way as well as extensive and transparent justification.

Depending on the treatment of logical remainders (if there are any) procedure no. 3 may come close to the parsimonious or to the complex solution. Another strategy,

- 4) consists of using only easy counterfactual configurations and thus producing intermediate solutions. This proposal was submitted by Ragin and Sonnett (2005). It builds a bridge between highly complex and overly parsimonious solutions. It thus resembles strategy no. 3. However, conditions are excluded “from the complex solution that are inconsistent with existing knowledge, while [...] the [...] solution constructed [...] must be a subset of the most parsimonious solution” (Ragin and Sonnett 2004:17). Whereas in strategy no. 3 the number of potential assumptions about logical remainders is decreased artificially by coding rows in the truth table, strategy no. 4 removes redundant conditions in the above mentioned manner. It thus contains only some of the logical remainders the most parsimonious solution comprises. Both modes (no. 3 and 4) are based on prior knowledge and theoretical reasoning.

There is no strategy that handles limited diversity in an optimal way. They all have their strengths and pitfalls. If one decides to get parsimonious, general solutions that can be interpreted in a meaningful manner (a seductive point) the inclusion of simplifying assumptions about logical remainders seems obvious. However, this strategy raises the proportion of not occurred combinations in relation to combinations with empirical evidence. The more simplifying assumptions are allowed for, the less explanatory conclusions will rest on empirical findings. The empirical foundation of a causal theory resting on such a QCA thus becomes more and more unsound. The potential number of simplifying assumptions on logical remainders rises excessively with the inclusion of each new condition. The formula for the number of unique logical combinations in csQCA clarifies that:

$$(1) \quad n = 2^k$$

with n as the number of unique logical combinations of factors and k the number of conditions. Each new condition doubles the number of unique logical combinations. For example, with seven conditions there are 128 combinations, with eight conditions the number of combinations is 256. The maximum of simplifying assumptions is different from the number of unique combinations of conditions:

$$(2) \quad z_{\max} = 2^k - 1$$

with  $z$  as the number of simplifying assumptions on logical remainders and  $k$  the number of conditions. The maximum occurs when there is empirical evidence for just one logical combination.

The other extreme of handling limited diversity is to ignore all simplifying assumptions about logical remainders and to admit only observed logical combinations to logical minimisation. All causal statements will rely on combinations which there is empirical evidence for. The drawback is that the QCA is likely to end in overly complex solution formulas – long and hard to interpret. This holds true all the more for mvQCA, as the number of different condition values is potentially higher. The worst case arises when the analysis reveals one complex causal path including numerous conditions for each and every single case – the findings cannot be interpreted or generalized with reasonable effort any more.

### **The fsQCA two-step approach**

A middle course between overly complex and oversimplified solution formulas would consist in an analysis that allows for simplifying assumptions about logical remainders *and* that gets a grip on the problem of limited diversity. By the introduction of a two-step approach Carsten Q. Schneider and Claudius Wagemann (2003; 2006; Schneider 2009) offered such an instrument. The authors' aim was to "show how this contributes to remedying the problems of limited diversity and to achieving digestible but, nevertheless, theoretically subtle results" (Schneider and Wagemann 2006:759).

The module rests upon the classification of conditions according to their remoteness relative to the outcome observed. Factors thus can be either causally remote or causally proximate. According to Schneider and Wagemann remote factors are relatively stable over time, remote in a spatiotemporal way, and out of a manipulative reach of the actors. They are regarded as given, structural constraints. Proximate factors in turn change easily over time, they are spatially and temporally close to the outcome and subject to manipulations of actors.

[Table 1 here]

It is not always easy to say properly whether a factor is a proximate or remote one. According to the authors the question of remoteness and proximity therefore has to be seen as a continuum whose poles are labelled according to both features. Depending on what someone examines the groups of remote and proximate factors may be entirely different from the groupings someone else makes, even if both make use of identical conditions. First and foremost, this is due to the focus on various outcomes: The question of remoteness of conditions basically depends on the question what is explained.

*Step one* comprises the enquiry of remote factors. The authors recommend the inclusion of simplifying assumptions on logical remainders. The results of this step will be outcome *enabling* conditions, contexts in some measure that need not meet high requirements. The required consistency values in this step should not be too rigorous.<sup>1</sup> In the article that

---

<sup>1</sup> In csQCA and mvQCA the consistency of sufficient conditions provides information about how many cases that display a certain value cause the outcome, too, in relation to all cases where the value is present. This factor ranges from 0 (no case that displays the value also displays the outcome) to 1 (all cases that exhibit the value

introduces the two-step approach Schneider and Wagemann bring in an empirical example. Accordingly, a configuration must capture at least one case and pass a consistency threshold of 0.7 to be included into the minimisation process. In his study on democratization in Europe and Latin America Schneider proposes an even more rigid value of 0.8 (Schneider 2009:81). Although this makes clear that no definite threshold exists, a consistency value somewhere between 0.7 and 0.8 seems reasonable. The consistency value of the achieved formula may be surprisingly low – but that does not pose a problem. It only purports that there are cases that display a certain context but not the outcome. What is more, the consistency of the formula will rise after the second step as this will specify the explanations.

The *second step* aims at identifying combinations of proximate factors that lead to the observed outcome within the contexts found in step one. The number of second step-analyses is thus identically equal to the number of contextual, i. e. remote, configurations: in each second-step analysis a group of cases covered by one context is examined separately from the cases covered by other contexts. The conditions standing in the centre of interest are the proximate ones. Whereas in their joint paper Schneider and Wagemann recommend the unmodified consistency value of 0.7 in step two, Schneider (Schneider, 2009, p. 85) uses a threshold of 0.9 in his study of democratization. He justifies this decision with higher requirements in order to achieve consistent solutions. Assumptions on logical remainders are excluded from the minimisation process in step two. The consequence: the most complex, though highly consistent solution.

The approach claims credit for two things which are indicative that it is applied regularly in a QCA: first, according to Schneider and Wagemann it gets a grip on a typical problem of QCA in general, and fsQCA in special: limited diversity. This merit results from the mere division of QCA into two separate steps:

$$(3) \quad z_{\max} = 2^{k_1} - 1 + 2^{k_2} - 1$$

with  $z$  as the maximum number of simplifying assumptions on logical remainders,  $k_1$  the number of remote conditions and  $k_2$  the number of proximate conditions. However, the formula is in need of correction:

$$(4) \quad z_{\max} = (2^{k_1} - c) + (2^{k_2+1} - 1) * c$$

where  $z$  is the maximum number of simplifying assumptions on logical remainders and  $k$  the number of conditions in the respective step.  $C$  is the number of elicited contexts.  $2^{k_1} - c$  is thus the maximum number of logical remainders in step one,  $2^{k_2+1} - 1$  is the maximum number of logical remainders of only *one analysis* in step two. It therefore has to be multiplied with  $c$ , the number of analyses that have to be made in step two. Thus, the effect of Schneider and Wagemann's two-step approach is not as big as it may seem at first glance (Fig. 1). The

---

exhibit the outcome, too) (cf. Schneider and Wagemann 2007:86-90). For instance a consistency of 0.8 says that 80 percent of cases that display one value of condition in question also cause the outcome whereas 20 percent do not. What matters for the consistency value of the sufficient condition  $X$  in a fsQCA is the sum of  $X$ -values that are higher than the corresponding  $Y$  (Schneider and Wagemann 2007:202-211).

number of contexts plays a major role in the determination of simplifying assumptions on logical remainders – it has been neglected before.

[Figure 1 about here]

An example will illustrate the chain of thought: An analysis of the outcome O with the help of four remote (A, B, C, D) and four proximate (E, F, G, H) factors illustrates – from a purely methodological perspective – that the maximum number of simplifying assumptions on logical remainders depends 1) on the number and distribution of factors in both causal groups, and 2) on the number of detected contexts. Assuming that all eight factors are dichotomous, that they are equally distributed in both groups, and that in the analysis of external factors (step one) three outcome-enhancing contexts have come apparent which rely on the maximum number of simplifying assumptions on logical remainders, i. e.  $2^{k_1} - 3 = 2^4 - 3 = 13$ , step two comprises three analyses: One with context one and all internal factors, one with context two and all internal factors, and one with context three and all internal factors. What has not been taken into account is the following: As it is not known whether some proximate conditions will or will not outplay the contexts the latter have to be included in the second step analyses. Thus the maximum number of simplifying assumptions on logical remainders in *one* second-step analysis is  $2^{k_2+1} - 1 = 2^5 - 1 = 31$ . This number has to be multiplied with the number of analyses  $c = 3$  that have to be made in step two. The maximum sum of assumptions in step two thus is 93, the maximum number in the whole analysis process is no less than 106 and not – as has been suggested – 30 (Schneider and Wagemann 2006:762). The situation changes with the number of determined contexts: *Ceteris paribus* the presence of only one context maximally implies 46 simplifying assumptions on logical remainders, the presence of two contexts comes along with maximally 76 simplifying assumptions on logical remainders, three contexts imply 106, four contexts imply 136 simplifying assumptions.

The proposition of a considerable impact of the two-step approach on limited diversity can only be maintained in three cases: 1) The number of identified contexts is low, or 2) logical remainders are excluded in step two (Schneider and Wagemann 2012: 254) or, 3), no minimisation process takes place in step two.

The second – and more important – merit of Schneider and Wagemann’s two-step approach is based on the fact that it coerces the researcher to consider both proximate and remote conditions. As a result the analysis accounts both for causal depth and causal mechanisms without pitting them against each other, as is done quite often in statistical analyses (Schneider and Wagemann 2006:761): A range of studies emphasizes the explanatory power of proximate factors as they generally explain a good share in variance of the explanandum compared to remote factors. Accordingly, remote factors often get left out. Yet the focus on proximate factors bears the risk of tautological explanations that disregard the long-term genesis of a certain outcome. Such explanations remain in shallowness and lack causal depth.<sup>2</sup> The other – though considerably less common – extreme consists of naming the structural and historical constraints of a certain outcome while at the same time concealing the proximate conditions that serve as a connecting link in the causal chain between remote contexts and the

---

<sup>2</sup> For example: “The quality of democracy depends on institutional arrangements such as precautionary measures against extremist groups”.

outcome. The effect of causal mechanisms connecting a remote and a proximate condition in order to trigger the outcome is neglected. Such propositions feature pivotal missing links.<sup>3</sup> Inquiries having recourse to the two-step approach in turn usually find a good balance between causally remote and proximate factors, as other students in social science have already done (See, for example Lehbruch 1979; Lipset and Rokkan 1967; Crouch 2003; Kitschelt 1999; Kitschelt 2003): The approach “offers a practical solution to the general need to contextualize causal statements and thus to formulate middle-range theories” (Schneider and Wagemann 2006:775-776). What therefore – in a nutshell – speaks for it is that it avoids tautologies (Schneider and Wagemann 2003: 17-18), missing links, and that it gets a grip on limited diversity in fsQCA (Schneider and Wagemann 2003: 18-20).

### The mvQCA two-step approach

Multi-value QCA advances the original csQCA by allowing not only for dichotomous, but also for multinomial conditions (Cronqvist 2007; Schneider and Wagemann 2007: 262-265). In doing so it maintains the information level of the original data to a higher degree than csQCA does. This, however, intensifies the problem of simplifying assumptions on logical remainders concomitant with limited diversity (Schneider and Wagemann 2012:260-263):

$$(5) z_{\max(\text{mvQCA})} = \prod_{i=1}^{k_2} a_i - 1$$

with  $z$  as the number of simplifying assumptions and  $a$  as the number of values condition 1, 2, 3, ...,  $i$  can take. For instance, a one-step mvQCA drawing on eight conditions of which four are dichotomous and four are trichotomous the maximum number of simplifying assumptions on logical remainders is 1295 (1296 – 1). The worst case scenario of a one-step csQCA application with eight conditions includes only one fifth of the number of logical remainders.

Unfortunately, the two-step procedure unfolded by Schneider and Wagemann has been proper for the use within fsQCA only (Schneider and Wagemann 2007:262). No appropriate instrument similar to it got a grip on limited diversity and causal imbalance in mvQCA – the first problem being even more severe in mvQCA than in csQCA. Therefore I aim at transferring the merits of Schneider and Wagemann’s fsQCA approach to the analysis of multinomial conditions. In order to do that functional equivalents in mvQCA need to be identified.

The division of factors according to their causal remoteness relative to the outcome is the centrepiece of the two-step approach. It avoids both causally tautological statements and statements that ignore the explanatory power of causal mechanisms. The two analysing steps hence are the elements that have to be maintained. Moreover, the first-step analysis is supposed to produce contexts in which the outcome occurs. The second step is supposed to specify the contextual explanation by introducing proximate factors.

Yet, apart from this the new approach differs from Schneider and Wagemann’s module. For reasons of clarity let me resort to the earlier example in which the outcome  $O$  will be

---

<sup>3</sup> For example Robert Putnam’s study that attributed the performance of Italian regional governments to medieval coinages without telling how both phenomena are causally linked to each other over centuries.

explained by the four remote conditions A, B, C, and D and by the four proximate conditions E, F, G, and H. Two conditions of each group are dichotomous, the rest is trichotomous; the 20 cases in the focus are displayed in the following configuration table.

[Table 2 here]

As in Schneider and Wagemann's approach the first step consists of determining contexts – however, here irrespective of the question whether they affect the outcome favourably or less favourably. This is a difference between the fsQCA and the mvQCA two-step approach which is caused by varying meanings of inconsistency and different set theory in fsQCA and mvQCA (Schneider and Wagemann 2012:127). They need some preliminary remarks at this point: “For a condition [in fsQCA] to be sufficient for Y, each case's membership in the condition must be equal to or smaller than its membership in Y.” (Schneider and Wagemann 2012:68) Inconsistency in the first-step fsQCA thus comes into being because there are cases that display higher membership values in the context than in the outcome.<sup>4</sup> This inconsistency can be resolved – partially or completely – by introducing proximate conditions, for which as many *inconsistent* cases as possible display lower membership values than in the outcome. In other words: In order to raise consistency in the fsQCA two-step approach not only new conditions must be introduced, but the inconsistent cases must feature lower membership values in them than in the outcome. However, the chances for a maximum of consistency through step two shrink with the number of inconsistent cases after step one. By setting the consistency threshold in the first step not too low the fsQCA two-step approach makes the assumption that a certain configuration of remote conditions already favours the outcome. In doing so the number of cases for which a “suitable” proximate condition must be found decreases whereas the chances for a maximum of final consistency increase.

In mvQCA the problem is quite different and unevenly easier to solve: Inconsistency and the presence of contradictory cases are two sides of the same coin: Resolving inconsistency is equivalent to resolving a contradiction. This can be achieved by adding conditions, by respecifying the definition of the population of interest or by respecifying the definition, the conceptualization, and/or the measurement of the outcome or the conditions (Cf. Schneider and Wagemann 2012:120-121). The first strategy will be deployed in the mvQCA two-step approach. As by the very inclusion of new, more precisely: proximate, conditions the consistency of the solution can be raised dramatically, the interim solution formula after step one does not need to meet any consistency requirements. Accordingly, the empirical analysis will reveal the causal meaning of a certain context: it could produce the outcome on its own (when a remote causal term is already fully consistent), it could enhance the outcome (when a remote causal term alone is highly, though not fully consistent) or it could be neutral and form a most similar case design at worst (when its consistency is low).

Bearing this in mind, the first-step analysis can begin. The following configuration table serves as its starting point. It only includes the remote factors A, B, C, and D.

---

<sup>4</sup> As the original consistency measure for sufficient conditions applies (Cf. Schneider and Wagemann 2006: footnote 30), it does not matter – as in the modified consistency measure – how far they inconsistent cases lie below the diagonal in the X-Y-plot.



[Table 3 here]

As the table illustrates, there is a contradictory line: In the cases i, n, o, p, r, and s one and the same configuration of remote conditions yields O(1), in case e it leads to O(0). Let us first ignore them and conduct the first-step analysis as if no contradictions existed. In this step simplifying assumptions can be justified with the fact that some contexts will already be fully consistent and thus not be in need of a second-step analysis inclusive of a minimisation process. The exclusion of simplifying assumptions in step one might thus result in overly complex final solution terms. What is more, the mere number of simplifying assumptions does not speak against their inclusion: The module reduces them dramatically. However, if the inclusion of simplifying assumptions discomforts the user, nothing speaks against their exclusion from a pure technical perspective.

In order to achieve a parsimonious interim solution simplifying assumptions on logical remainders are allowed in the minimisation process. Contradictions and cases with missing outcomes are excluded. Accordingly, the partial interim solution  $A(0)D(1) + A(1)B(0) \rightarrow O(1)$  is based on twelve simplifying assumptions: The first term covers case f, the second one covers a and l, both are fully consistent as they represent no contradictory cases.

However, the cases displaying O(1) and being part of the contradictory line at the same time are left unexplained yet. Let us now turn to these problem cases: One might argue that an equilibrium of cases featuring and not featuring the outcome within one configuration table row or even an imbalance to the detriment of the relevant outcome (see the configuration table) speaks for a theoretically questionable selection of remote conditions. However, the opposite might be true: It could indicate the sensitivity of the outcome to the interaction of varying proximate factors with similar contexts. For instance, one could state that the stability of democracy (outcome) in ethnically heterogeneous countries (context) might be due to the type of democracy institutionalised in the country, such as a more majoritarian or more consensus based democracy (proximate factors). Thus an analysis of the cases in which one and the same context leads to different results reflects a most similar case design.

First of all, the contradictory cases should be separated from the remaining cases: Contradictions are set to “explain”, whereas “outcome 0”, “outcome 1”, and “missing outcomes” are excluded from the minimisation process. Basing on five simplifying assumptions the solution term for the contradictory row is:  $B(0)D(0) \rightarrow C$ . The combination of B(0) and D(0) thus leads to a contradictory line in the case set. But how can this be interpreted regarding the outcome? To put it briefly: That depends on the outcome that is explained. Due to a proportion of 6:1 in the row the consistency of this term regarding outcome O(1) is 85.7 per cent, regarding O(0) it is  $100 - 85.7 = 14.3$  per cent. Strictly speaking, this causal term is no empirical proposition any more as it admits two contradicting events. Let us, however, recall that this term merely shapes a homogeneous background for a certain group of cases. Therefore, it is actually no causal term at all. This must be accepted for the time being. The interim solution term after step one is:  $A(0)D(1) + A(1)B(0) + B(0)D(0) \rightarrow O(1)$ . Whereas the first two terms are fully consistent, the last one covers the contradictory case group. The coverage of the formula equals 100 per cent as all cases that display the outcome O(1) are explained. The interim solution thus represents a causal theory of maximum range, however with substantial inconsistencies.

The following second-step analyses add the proximate conditions E, F, G, and H to the three contexts in order to eradicate the blur of inconsistency. Even though the first two terms of the interim solution do not call for the addition of proximate factors (and thus a second-step analysis) as they are fully consistent, I advocate counterfactual analyses in order to find out whether certain proximate factors may contribute to the comprehension of causal logics. They are called “easy” counterfactuals (Ragin and Sonnett 2005): It is easily conceivable that some cases which are consistently explained by one and the same context also share the identical internal factors which could be crucial in understanding the occurrence of the outcome. For instance, theoretical reasoning and an in-depth analysis of the relevant cases displaying outcome O(1) might reveal that it is the conjuncture of the remote conditions A(1) and B(0) with the proximate factor G(0) that yields O(1). The incorporation of G(0) does not make sense from a methodological perspective because it is logically redundant, but it might be useful from a causal interpretative perspective as it helps to track the nascence of the outcome.

An explanation drawing on such easy counterfactuals would consider both causal depth (by the inclusion of remote conditions) and causal mechanisms (by the consideration of proximate conditions) without at the same time making additional assumptions on logical remainders. The causal propositions resulting from such a two-step QCA are grounded to a high degree on empirical evidence.

The second-step analyses are important for another reason: The consistency of the solution term B(0)D(0) has to be raised. For this purpose the contradictory case group is separated from the data set. The following analysis focuses on the seven cases.

[Table 4 here]

As the proximate conditions are only to specify the interim solution B(0)D(0) in the way that they show why – *ceteris paribus* – in the case of e it comes to the outcome O(0) whereas six other cases display O(1) all remote factors can be omitted from this step. This is accounted for by a different function logic of mvQCA compared to fsQCA. The second-step analysis includes only four conditions.

Although Schneider and Wagemann advocate a rigorous handling of simplifying assumptions on logical remainders in step two I think the choice should be left to the user: The module allows both the exclusion of such assumptions and thus a more complex solution and the inclusion of simplifying assumptions on logical remainders and thus the most parsimonious solution. Anyway, their number will be low due to the split of the analysis into two steps – therefore, this argument cannot be urged anymore.

By the inclusion of logical remainders QCA produces one solution term with the help of the four proximate conditions:  $F(0) \rightarrow O(1)$ . The term covers all cases (i, n, o, p, r, and s) that feature the outcome O(1). This formula is based on one single simplifying assumption. As this term only refers to the seven contradictory cases the formula does not reflect the causal logic of the entire case set: First, this term has to be connected with the context that applies to all seven cases. This is achieved by the help of a logical multiplication sign. The causal term

explaining all seven cases incorporated in step two hence is:  $B(0)D(0) * F(0) \rightarrow O(1)$ .<sup>5</sup> In words: In the context of B(0) and D(0) the proximate factor F(0) will result in O(1).

Second, the complete solution formula – including the factor G(0) that is elicited by the counterfactual analysis – has to be compiled by the help of logical operators:  $A(0)D(1) + A(1)B(0) * G(0) + B(0)D(0) * F(0) \rightarrow O(1)$ . The outcome O(1) thus comes up in the context A(0)D(1) or when G(0) occurs in the context A(1)B(0) or when F(0) takes place in the context B(0)D(0). Though this maximally consistent, high-coverage solution formula is the most parsimonious one it only rests on 18 simplifying assumptions. A maximally parsimonious one-step solution would be even more parsimonious<sup>6</sup> – however it would not have illuminated the effectiveness of three contexts and the causal mechanisms between two of them and certain proximate conditions. What is more, it would have rested on 318 simplifying assumptions on logical remainders. Its empirical basis is thus slim compared to the result of this two-step analysis.

An analogous procedure follows for the analysis of the outcome O(0) – with the difference that the contradictory line (i, n, o, p, r, s) does not need to be separated from the remainder of the case set by an individual analysis. This has already been carried out in the analysis of O(1). The interim solution is:  $A(0)D(2) + A(1)B(1) + B(1)C(0,1) + B(0)D(0) \rightarrow O(0)$ . The formula features a coverage of 100 per cent, all terms – except the last one (0.14) – are fully consistent. The second-step analysis of the inconsistent term leads to the formula  $F(1) \rightarrow O(0)$ . The complete solution formula resulting from this two-step analysis is:  $A(0)D(2) + A(1)B(1) + B(1)C(0,1) + B(0)D(0) * F(1) \rightarrow O(0)$ . The beneficent effect of the two-step approach proves itself here, too: The number of simplifying assumptions in a one-step mvQCA would have amounted to no less than 784 – compared to 14 simplifying assumptions in the two-step mvQCA. Additionally, the causal mechanisms within the last term would have been disregarded.

## Summary

MvQCA is a variant form of QCA that is not untainted by the problem of limited diversity, in general, and the problem of simplifying assumptions on logical remainders, in special. The resulting problem is that causal statements may lack empirical evidence and thereby not only become simple, but simplistic. Beyond that, the insufficiency of shallowness or missing causal links applies to propositions resulting from mvQCA, too.

The two-step approach – originally developed for fsQCA – proved to be a helpful tool for mvQCA. For this purpose functional equivalents had to be found that are applicable to the analysis of multinomial conditions: the split of factors according to their causal remoteness and the subsequent analysis in two steps. As the fsQCA module it finds “the right balance between the two core features: causal depth and causal mechanisms” (Schneider and Wagemann 2012:254). It is especially the interaction of remote and proximate factors that the mvQCA two-step approach sets into focus: Although the first-step analysis might reveal fully consistent and thus complete causal explanations for the emergence of an outcome virtually

---

<sup>5</sup> The multiplication sign can be omitted, However, its presence makes the causal mechanism between the contexts and the proximate conditions clear.

<sup>6</sup>  $D(0)F(0) + E(0)F(1)G(0) \rightarrow O(1)$ .

nothing speaks against a second-step counterfactual analysis that adds logically redundant, but causally crucial proximate factors that all cases represented by one and the same context share. The inclusion of such conditions does not come at the cost of more simplifying assumptions on logical remainders (and thereby of a thinner empirical foundation). Instead, only an in-depth analysis of the few relevant cases may unveil them. Thereby, QCA's aim of oscillating between theory and empiricism, between qualitative and comparative claims is achieved.

What is more, by conducting an autonomous analysis of contradictory lines in the first step analysis of the outcome O(1) and O(0) the interaction of certain contexts and varying proximate conditions is enlightened. The procedure helps to identify causal mechanisms in most similar cases. For instance, in the hypothetical example the similar context of B(0)D(0) leads to O(0) an O(1) all at once – at least so it seemed at first sight. It is, however, the proximate condition F that matters. If it assumes the shape of F(0), it results in O(1), if F(1) comes into being, this will lead to O(0) – a perfect example of causal mechanisms the mvQCA two-step approach brings to light.

A considerable side benefit of the module is the reduction of simplifying assumptions on logical remainders compared to a one-step mvQCA by the mere split into two steps: There are three formulas for the maximum number assumptions about logical remainders depending on the result of the first step. The equation for an analysis revealing both fully consistent contexts and contradictory lines reads as follows:

$$(6a) z_{\max 2S} = \prod_{i=1}^{k_1} a_i - 1 + \prod_{i=1}^{k_1} b_i - 1 + \prod_{i=1}^{k_2} c_i - 1$$

The maximum number of assumptions about logical remainders in analyses that reveal only fully consistent cases the equation is:

$$(6b) z_{\max 2S} = \prod_{i=1}^{k_1} a_i - 1$$

The maximum number of assumptions about logical remainders in analyses that reveal only contradictory lines in the first step the equation is:

$$(6c) z_{\max 2S} = \prod_{i=1}^{k_1} b_i - 1 + \prod_{i=1}^{k_2} c_i - 1$$

- with  $z_{\max 2S}$  as the maximum number of logical remainders in the two-step approach,
- a the number of possible values of the i-th remote condition in the analysis of consistent contexts,  $k_1$  the number of remote conditions,
- b the number of possible values of the i-th remote condition in the analysis of contradictory cases
- c the number of possible values of the i-th proximate condition,  $k_2$  the number of proximate conditions.

The appropriate equation for the maximum number of simplifying assumptions about logical remainders thus can be read off from the configuration table. In the example it is calculated according to equation (6a) as both fully consistent contexts and contradictory lines appear.

The number is:  $36 - 1 + 36 - 1 + 36 - 1 = 105$ . In case of only consistent lines (6b) it would have been 35 (6b); in the case of contradictory cases only (6c) it would have amounted to no more than 70.

No matter which equation applies – the number of simplifying assumptions about logical remainders is decreased dramatically compared to a one-step mvQCA. Please note that this number may also rest on the use of so called incoherent counterfactuals – identical simplifying assumptions in the analysis of consistent contexts and the analysis of contradictory rows – or even on implausible counterfactuals (Schneider and Wagemann 2012:198-199). Their exclusion will contribute to the further reduction of simplifying assumptions and thus the stabilisation of the empirical foundation of the causal propositions. This, however, is a question that touches another area of QCA enhancements.

## References

Cronqvist, Lasse. 2007. *Konfigurationelle Analyse mit Multi-Value QCA als Methode der Vergleichenden Politikwissenschaft mit einem Fallbeispiel aus der Vergleichenden Parteienforschung (Erfolg Grüner Parteien in den achtziger Jahren)*. Inauguraldissertation. Marburg.

Crouch, Colin. 2003. *Institutions within which real actors innovate*. in: Mayntz, Renate, and Fritz Scharpf, eds. *Die Reformierbarkeit der Demokratie. Innovationen und Blockaden*. Frankfurt/Main: Campus-Verlag.

Kitschelt, Herbert. 1999. *Accounting for Outcomes of Post-Communist Regime Change. Causal Depth or Shallowness in Rival Explanations*. Duke University. Atlanta.

Kitschelt, Herbert. 2003. *Accounting for Postcommunist Regime Diversity. What Counts as a Good Cause?* in: Ekiert, Grzegorz, and Stephen E. Hanson, eds. *Capitalism and Democracy in Central and Eastern Europe. Assessing the Legacy of Communist Rule*. Cambridge: Cambridge University Press: 49-86.

Lehmbruch, Gerhard. 1979. *Liberal corporatism and party government*. In: Schmitter, Philippe C., and Gerhard Lehmbruch, eds. *Trends toward corporatist intermediation*. Beverly Hills, CA.: Sage.

Lipset, Seymour Martin, and Stein Rokkan. 1967. *Party systems and voter alignments: Cross-national perspectives*. New York: Free Press.

Mannewitz, Tom: "Two-Level Theories in QCA. 2011. A Discussion of Schneider and Wagemann's Two-Step Approach." *Compass Working Paper Series, 2011-64*. Retrieved April 19, 2011 (<http://www.compass.org/pages/resources/compasswp.html>).

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

Ragin, Charles C. 2000. *Fuzzy-Set Social Science*. Chicago and London: The University of Chicago Press.

Ragin, Charles C., and John Sonnett. 2005. *Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases, and Comparative Analysis*. in: Kropp, Sabine and Michael Minkenberg, eds. *Vergleichen in der Politikwissenschaft*. Wiesbaden: Verlag für Sozialwissenschaften:180-97.

Rihoux, Benoît, and Charles C. Ragin, eds. 2009. *Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and Related Techniques*. Thousand Oaks: Sage Publications.

Sager, Fritz, and Céline Andereggen. 2012. „Dealing with Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis.” *American Journal of Evaluation*, Vol. 33:60-78.

Schneider, Carsten Q. 2009. *The Consolidation of Democracy. Comparing Europe and Latin America*. New York: Routledge.

Schneider, Carsten Q., and Claudius Wagemann. 2003. “Improving Inference with a ‘Two-step’ Approach: Theory and Limited Diversity in fs/QCA.” *EUI Working Paper SPS 2003/7*.

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*, Vol. 45:751–786.

Schneider, Carsten Q., and Claudius Wagemann. 2007. *Qualitative Comparative Analysis und Fuzzy Sets. Ein Lehrbuch für Anwender und jene, die es werden wollen*. Opladen & Farmington Hills: Barbara Budrich.

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

Vink, Maarten, and Olaf van Vliet. 2007. “Not quite crisp, not yet fuzzy? Assessing the potentials and pitfalls of multi-value QCA.” *Compass Working Paper Series WP 2007-52*.

## Tables

Table 1: Remote and proximate factors according to Schneider and Wagemann

| Remote Factors  | Proximate Factors  |
|---|--|
| <ul style="list-style-type: none"> <li>• spatiotemporally distant to the outcome</li> <li>• stable over time</li> <li>• out of manipulative reach of the actors involved</li> </ul> | <ul style="list-style-type: none"> <li>• spatiotemporally close to the outcome</li> <li>• vary easily over time</li> <li>• can be manipulated by actors</li> </ul> |

Source: Own compilation.

Table 2: Configuration table (all cases, all conditions)

|     |    |     |      |    |    |    |    |   |       |
|-----|----|-----|------|----|----|----|----|---|-------|
| v1: | A  | v2: | B    |    |    |    |    |   |       |
| v3: | C  | v4: | D    |    |    |    |    |   |       |
| v5: | E  | v6: | F    |    |    |    |    |   |       |
| v7: | G  | v8: | H    |    |    |    |    |   |       |
| O:  | O  | id: | Case |    |    |    |    |   |       |
| v1  | v2 | v3  | v4   | v5 | v6 | v7 | v8 | O | id    |
| 1   | 0  | 2   | 1    | 0  | 1  | 0  | 0  | 1 | a     |
| 0   | 0  | 1   | 2    | 0  | 1  | 2  | 1  | 0 | b,k   |
| 0   | 1  | 1   | 2    | 0  | 1  | 2  | 0  | 0 | c     |
| 0   | 1  | 0   | 0    | 0  | 1  | 2  | 0  | 0 | d     |
| 0   | 0  | 0   | 0    | 1  | 1  | 0  | 0  | 0 | e     |
| 0   | 1  | 2   | 1    | 0  | 1  | 0  | 2  | 1 | f     |
| 0   | 1  | 2   | 2    | 0  | 0  | 2  | 1  | 0 | g     |
| 0   | 0  | 0   | 2    | 0  | 0  | 0  | 0  | 0 | h     |
| 0   | 0  | 0   | 0    | 1  | 0  | 1  | 0  | 1 | i,n   |
| 1   | 1  | 2   | 1    | 0  | 0  | 0  | 2  | 0 | j,m   |
| 1   | 0  | 1   | 2    | 0  | 1  | 0  | 0  | 1 | l     |
| 0   | 0  | 0   | 0    | 0  | 0  | 1  | 0  | 1 | o     |
| 0   | 0  | 0   | 0    | 1  | 0  | 0  | 0  | 1 | p,r,s |
| 0   | 1  | 2   | 2    | 1  | 1  | 0  | 1  | 0 | q     |
| 0   | 1  | 2   | 2    | 0  | 1  | 2  | 1  | 0 | t     |

Source: Own compilation.

Table 3: Configuration table (all cases, remote conditions only)

|     |    |     |      |   |               |  |  |
|-----|----|-----|------|---|---------------|--|--|
| v1: | A  | v2: | B    |   |               |  |  |
| v3: | C  | v4: | D    |   |               |  |  |
| O:  | O  | id: | Case |   |               |  |  |
| v1  | v2 | v3  | v4   | O | id            |  |  |
| 1   | 0  | 2   | 1    | 1 | a             |  |  |
| 0   | 0  | 1   | 2    | 0 | b,k           |  |  |
| 0   | 1  | 1   | 2    | 0 | c             |  |  |
| 0   | 1  | 0   | 0    | 0 | d             |  |  |
| 0   | 0  | 0   | 0    | C | e,j,n,o,p,r,s |  |  |
| 0   | 1  | 2   | 1    | 1 | f             |  |  |
| 0   | 1  | 2   | 2    | 0 | g,q,t         |  |  |
| 0   | 0  | 0   | 2    | 0 | h             |  |  |
| 1   | 1  | 2   | 1    | 0 | j,m           |  |  |
| 1   | 0  | 1   | 2    | 1 | l             |  |  |

Source: Own compilation.



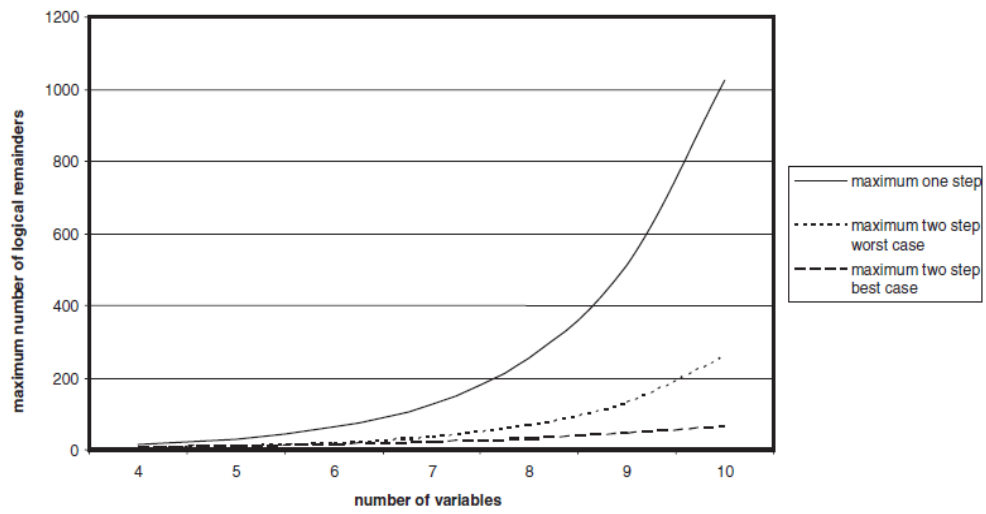
Table 4: Contradictory case group

| v1: | E  | v2: | F    |   |       |
|-----|----|-----|------|---|-------|
| v3: | G  | v4: | H    |   |       |
| O:  | O  | id: | Case |   |       |
| v1  | v2 | v3  | v4   | O | id    |
| 1   | 1  | 0   | 0    | 0 | e     |
| 1   | 0  | 1   | 0    | 1 | i,n   |
| 0   | 0  | 1   | 0    | 1 | o     |
| 1   | 0  | 0   | 0    | 1 | p,r,s |

Source: Own compilation.

## Figures

Figure 1: The alleged effect of the two-step approach on the maximum number of logical remainders



Source: Schneider, Wagemann 2006:763.