



Limited diversity and QCA solution types: assumptions and their consequences

Judith Glaesser¹ 

Accepted: 23 June 2022 / Published online: 8 September 2022
© The Author(s) 2022

Abstract

Limited diversity is a term employed in the context of Ragin’s Qualitative Comparative Analysis (QCA), but it describes a phenomenon which is widespread in social contexts: cases are usually not distributed evenly across all the possible combinations of factors linked to some outcome. Instead, they are often clustered together. To deal with limited diversity, the three solution types parsimonious, intermediate, and complex (or conservative) solution have been proposed. There is an ongoing debate as to the merits of each solution type. This paper contributes to the debate by focussing on what the implications of choosing each solution type are. In making this choice, researchers have to make certain assumptions, and the paper discusses what these are and how they vary depending on which solution type is being implemented, drawing on invented examples and examples from published work to bring out the consequences of these assumptions. It concludes that it is not obvious that any one solution type is superior, certainly not to the degree that the others always have to be ruled out. They rely on different kinds of assumptions and models of causation. Thus, depending on the research situation, it may be helpful to analyse different scenarios, including one(s) where we assume that our initial assumptions are wrong. Ideally, researchers are able to take steps to reduce or eliminate limited diversity since none of the existing solution types are without problems. However, since this is not always possible, it is best to be aware of the consequences of choices.

Keywords Qualitative Comparative Analysis (QCA) · Limited diversity · Solution types · Assumptions · Researcher choice · Causal model

✉ Judith Glaesser
Judith.glaesser@uni-tuebingen.de

¹ Methods Center, Tübingen University, Haußerstr. 11, 72076 Tübingen, Germany

1 Introduction

Limited diversity is a term employed in the context of Ragin's (1987, 2000, 2008) Qualitative Comparative Analysis (QCA), but it describes a phenomenon which is widespread in social contexts: cases are usually not distributed evenly across all the possible combinations of factors linked to some outcome. Instead, they are often clustered together. For example, in examining the roles of high parental class and high parental educational status for their children's educational outcomes, we will find that there are many cases where parental class and education are either both high or both low, but far fewer cases where one is high and the other low. This is not a sampling problem, but a reflection of the social structure: since class and education are interrelated, they will tend to co-occur in parents. Depending on the particular research focus and the sample under study, it is not unusual for some combinations not to have any empirical instances at all, and this is the phenomenon termed limited diversity by Ragin. One of the great advantages of using QCA is that this is brought out clearly via the truth table which sets out the distribution of cases before Boolean solutions are derived from it. Thus, unlike with other methods, the researcher is forced to focus his/her attention on how to deal with limited diversity at an early stage in the analytic process¹.

Kogut and Ragin (2006) describe limited diversity thus: "...nature rarely runs all experiments" (p.49), and they go on to note "This possibility raises the question of what should be the inference from missing configurations. The Boolean approach, unlike the central tendency of statistical analysis, forces the researcher to analyze the implications of unobserved logical combinations." To do so, Ragin proposes three solution types which deal with limited diversity in different ways: the complex (or conservative) solution, the intermediate solution, and the parsimonious solution (Ragin 2008, chaps. 8 and 9; Ragin and Sonnett 2005). There is an ongoing debate as to which solution type is best, with a wide range of analytic approaches examining the issues in different ways and coming to different conclusions (e.g., Baumgartner 2015, 2021; Baumgartner and Thiem 2020; Cooper and Glaesser 2016; Duša 2019; Haesebrouck 2022; Ragin and Sonnett 2005; Schneider and Wagemann 2012; Thiem 2019; Thomann and Maggetti 2020). Most recently, this debate was conducted in a special issue of *Quality & Quantity*, edited by Haesebrouck and Thomann (2021). Part of the disagreement as to which solution type is to be preferred presumably stems from what the aim of the analysis is taken to be, and how causation is to be established, i.e. what counts as evidence for causation. Causes may be conceived of as INUS conditions (Ragin and Sonnett 2005), i.e. as Boolean difference-makers (Baumgartner 2015). Alternatively, the focus may be on a theoretical understanding of causal processes and causal mechanisms. Such a focus is adopted, for example, in undertaking the kind of counterfactual reasoning which is an essential strategy in identifying easy counterfactuals for obtaining the intermediate solution (Ragin 2008, chaps. 8 and 9; Ragin and Sonnett 2005). The implications of differing research aims for the choice of solution type are summarised by Baumgartner: "One topical anchor point of this special issue [of *Quality & Quantity*] is the question which of QCA's parsimonious, intermediate, or conservative solution types should be produced and interpreted primarily. The answer depends heavily on what the search target of QCA is taken to be." (Baumgartner 2021, p.1).

¹ While it is possible, of course, to examine the relationships of factors amongst themselves with other methods, this is not an integral part of the analysis. Thus, researchers may not be aware of the extent of limited diversity in their data, and accordingly, they cannot take account of it in their analyses.

Table 1 Simple example of limited diversity

Strong Unions (U)	Strong Left Parties (L)	Generous Welfare State (G)	N of Cases
Yes	Yes	Yes	6
Yes	No	No	8
No	No	No	5
No	Yes	?	0

My own aim in this paper is to contribute to the debate by focussing on what the implications of choosing each solution type are. In making this choice, researchers have to make certain assumptions, and I will discuss what these are and how they vary depending on which solution type is being implemented. Drawing on invented examples and examples from published work, I intend to bring out the consequences of these assumptions. Thus, my aim is not to argue in favour of any particular solution type or theory of causation, since these depend on the researcher's philosophical position, the aims of the research and the specific situation (i.e., topic, availability of data, etc.). Instead, I intend to demonstrate in what way the choice of solution type is linked to the particular research aim and research topic, and I discuss what the possible implications of the choice are by spelling out the assumptions underpinning each solution type and their likely consequences.

2 Limited diversity and solution types

2.1 Limited diversity

Ragin (2008; Ragin and Sonnett 2005) uses an example with just two conditions to introduce limited diversity. The conditions are strong left parties and strong unions, and the outcome is generous welfare state. Table 1 reproduces the truth table (Ragin 2008, p.148).

The critical row is the last row, known as a remainder row, which has no cases, i.e. there are no countries where the unions are not strong but which have strong left parties. Therefore, it is impossible, on the basis of the data alone, to decide whether a generous welfare state is the result of strong left parties on their own, regardless of the strength of unions (L \rightarrow G, a parsimonious model), or whether it is the result of the strong left parties combined with strong unions (U*L \rightarrow G, a complex model). A researcher is thus left with the choice either of presenting both as explanatory models – a somewhat unsatisfactory position – or of making an assumption, based on counterfactual, theoretical reasoning, regarding what the outcome would be in countries where there are no strong unions combined with left parties. Depending on which assumption is made, the resulting explanatory model would be either the parsimonious or the complex one. A third possibility would be to prefer a parsimonious model regardless of any counterfactual reasoning and case knowledge on the grounds that the type of causal dependencies sought by Boolean methods are difference-makers, and that only parsimonious models are suitable for consistently identifying Boolean difference-makers (Baumgartner 2015).

Before I comment further on possible solution types, I will briefly discuss possible reasons for limited diversity. In Ragin's (fictitious but plausible) example, the lack of cases may have arisen because union movements precede the establishment of left parties and so there are no countries where there are strong left parties but no strong unions, since the lack of

strong unions has prevented the formation of strong left parties. The reason for the limited diversity would then be that the conditions are causally related to each other in a particular way, as well as to the outcome. Another example of such a situation is the one I gave at the beginning, where high parental education is a prerequisite for high parental class (and each is a putative causal condition for the outcome, children's educational status), making it more likely that the two co-occur. Another reason for limited diversity is that the co-occurrence of two causal conditions is logically impossible, for example when level of school qualifications and of vocational qualifications are both examined as conditions for some outcome such as social class. If having at least a basic type of school qualification is a prerequisite for being able to enter vocational training, then the combination of no school qualification and having a vocational qualification is not possible². Finally, it is possible that our sample is just too small for all the possible combinations to occur, even if they are not related amongst themselves and it is not logically impossible for them to co-occur. This situation of course is more likely to arise the smaller the sample and the larger the number of conditions under study (recall that the number of possible combinations of conditions rises exponentially, with 2^k giving the number of possible ways of combining k binary conditions).

It is important to be aware of the likely reasons for the limited diversity, since this has implications for the choice of strategy for dealing with it, as I discuss in Sect. 3.

2.2 Solution types

I do not explain the three solution types in detail here since this is beyond the scope of the present paper (for readers wishing to (re-)familiarise themselves with the procedures, see Ragin 2008; Ragin and Sonnett 2005). Instead, I comment on some key features and prerequisites of each solution type. They are each based on how the remainder row is treated. To obtain a complex (or conservative) solution, it is assumed that none of the missing configurations would have the outcome even if they could logically exist and if cases to populate them could be found. To obtain a parsimonious solution, remainder rows can be assumed to have whichever outcome helps produce a simpler solution. This may result in a dilemma: some of the assumptions necessary to produce the solution may be unrealistic or even demonstrably wrong. However, Baumgartner (2015) shows that the dilemma arises because of QCA's reliance on the Quine-McCluskey algorithm which had been developed in a different context. Baumgartner's own method, coincidence analysis (CNA), builds on QCA but uses a different, custom-built algorithm which dispenses with the need for assumptions altogether to produce parsimonious solutions³. Ragin has also proposed a third solution type, the intermediate solution where remainder rows may be assigned the outcome value thought to be plausible depending on assumptions based on theoretical reasoning and an understanding of the mechanisms. These rows can then be incorporated into the solution.

² Zhang's example reported in Duşa (2019), where two professors (A and B) have to agree that a candidate passes an exam (outcome C) seems to me to be related: $A * b \rightarrow C$ and $a * B \rightarrow C$ are logical impossibilities given the rules on the exam. In addition, the setup contains a causal chain: A and B cannot differ because of B's deference towards A. Accordingly, only some very specific lessons regarding how to deal with limited diversity may be drawn from this particular example, in my view.

³ The solutions produced by CNA are frequently identical to QCA's parsimonious solutions. However, as Swiateczak (2021) demonstrates, this is not always the case with noisy data, i.e. in a situation which is common in social research. The difference may be at least partly due to the different ways of handling consistency thresholds in CNA and QCA.

A key property of the three solution types is that they are subsets and supersets of each other, with the parsimonious solution being a superset of the intermediate solution, and this in turn a superset of the complex solution. Rather than one being superior to the others, Ragin suggests that they be viewed “as endpoints of a single continuum of possible results. One end of the continuum privileges complexity (no counterfactual cases allowed); the other end privileges parsimony (easy and difficult counterfactual cases are both allowed). Both endpoints are rooted in evidence; they differ in their tolerance for the incorporation of counterfactual cases.” (Ragin 2008, p.163/164) Thus, all three solution types are reflections of the constellation of conditions, but they differ in the level of detail at which the combinations linked to the outcome are described. By that, I mean that, while the complex solution describes all the combinations of factors empirically linked to the outcome – and only those – the parsimonious solution also describes cases empirically linked to the outcome, but omits some of the factors characterising these cases, while at the same time being broad enough to cover types of cases which have not been observed empirically. For example, the complex solution derived from Ragin’s example $L*U \rightarrow G$ describes the first truth table row only, while the parsimonious solution $L \rightarrow G$ describes the first truth table row and also covers the remainder row $L*u$. But since the parsimonious solution does not include U , it leaves out some detail pertaining to the empirically observed first row.

3 Assumptions and their consequences

In this section, I discuss assumptions made in the context of limited diversity and their possible consequences within different scenarios. The first is a research situation where all possible cases are already included in the analysis, i.e. the sample is the universe. For example, when an explanation of historical events is sought, the analysis concerns an outcome in the past, so that no new cases can possibly emerge to which the explanation may be applied, and the scope of the analysis concerns a known and limited number of cases. The second scenario is one where in principle more cases exist within the scope of the analysis.

In many research situations, it is clear which of the two scenarios we find ourselves in. If not, i.e. if we do not know whether more cases within the scope of the analysis might exist, which scenario we assume has implications for the choice of solution type and the counterfactual assumptions. One of the issues here is generalisation: if there is no intention to generalise from the findings to other cases outside the analysis, then it is possible to proceed as if scenario 1 applied, even if, in principle, more cases might be found. However, this is not without problems, since either the researchers themselves or those who draw on their research may apply any causal explanation derived on the basis of a particular sample to other cases of the same type, contrary to the original intention. But this is only warranted if the original sample was not biased.

3.1 Scenario 1: the sample of cases is the universe

In this section, I discuss research situations where the sample of cases is the universe, in other words, no other cases exist apart from those already included in the study. However, it is possible that this is itself an assumption because it is not always clear-cut whether more

cases could exist or not. Thus, throughout this section, I also discuss the implications of assuming that the study cases are the universe.

If no other cases apart from the ones providing the data for the study could possibly exist, then assumptions regarding the outcomes of those configurations without cases are not appropriate. In such a scenario, all three solution types produce correct models in the sense that they encompass causally relevant factors, and theoretical considerations and/or researcher preference concerning causal models determine which solution is adopted as the explanatory model. If we knew that no more cases could exist in Ragin's (Ragin 2008; Ragin and Sonnett 2005) example of conditions linked to the presence of a generous welfare state, we would draw on knowledge of the countries and of relevant mechanisms to help us decide whether strong left parties on their own are responsible for establishing a strong welfare state (and the presence of strong unions merely happens to coincide with strong left parties), or whether the unions had their own role in bringing this about in conjunction with the strong left parties' influence. If the former, then the parsimonious model may be adopted, if the latter, the complex model is appropriate. Both models are correct in the sense that they describe the situation, but they vary in their level of detail, and one may make more theoretical sense than the other.

Staying with this example, it may also be that it is *not* possible to decide that no more cases can exist, though if we know about the reason(s) why diversity is limited in the first place, we may well decide that certain *types* of cases could not exist. More cases could exist because we are studying a sample, not the population of cases, and they are either contemporary or, if historical, data exist in principle to examine additional cases within the scope of the analysis. Furthermore, the missing combination is not logically impossible. But if we assume or know that strong unions are a prerequisite for strong left parties – a causal claim in its own right – then we would never find cases where strong left parties coincide with the absence of strong unions, i.e. the remainder row in Table 1. (The row combining strong unions with the absence of strong left parties would not contradict this, since we would assume that strong unions are necessary but not sufficient for strong left parties.)

Arguably, the study by Berg-Schlosser and De Meur referred to by Haesebrouck (2022) is one where the study cases are the universe. They analyse the breakdown of democracy during the interwar period in Europe, and of course the interwar years were a specific period in time which is firmly in the past, so that no new cases can be added to the existing ones. Unless we assume that historical context does not matter and the same processes will occur at any point in time, we do not – indeed, should not – make assumptions regarding remainder rows since these will never be populated. The choice of model is then, again, informed by theoretical considerations and case knowledge. But the theoretical considerations do not concern what would happen in cases with the missing configurations, if they were to be found. Instead – since we know they do not exist – they concern the causal mechanism behind the outcome and which solution type better reflects this.

Clearly, in many cases the claim that no more cases exist is itself an assumption and/or a causal claim. The analysis of the breakdown of democracy during the interwar years may seem a straightforward case of the sample being the universe, but even here it is possible to argue that the causal processes would be similar during different historical periods. This means that other cases could exist and that generalisation from the study of the interwar years to other historical periods is possible. Another situation is one like Ragin's example of strong left parties and strong unions. The causal claim that there is some relationship

between strong left parties and strong unions is itself subject to scrutiny and may not be warranted, so that cases which are claimed not to exist on the basis of this causal claim do in fact exist and could be used to resolve the limited diversity⁴.

In the case of logically impossible combinations, the complex solution may be correct but silly. To refer to the example frequently drawn on by Schneider and Wagemann (2012), that of the (non-existent) pregnant man, a complex solution would include the combination pregnant*MALE, but the parsimonious solution including simply the condition MALE would be preferred. The complex solution, referring to men who are not pregnant, is not incorrect as such, but contains superfluous information. Obviously, this will not always be as apparent as in this example. Whichever one is chosen, it is worth bearing in mind that, as long there really are no more cases, neither the complex nor the parsimonious solution would contain elements which do not actually have the outcome. So from that point of view, they are both correct, though one may be preferred over the other either because it fits better with theoretical knowledge, or with case knowledge, and/or because the researcher conceives of causes as Boolean difference-makes, in which case only the parsimonious solution would be deemed correct.

3.2 Scenario 2: cases outside the analysis exist

A more common scenario than that described in the previous section is that we analyse a sample of cases and intend – implicitly or explicitly – to generalise the findings to other cases outside the sample. Since such cases might or do exist, it is important to proceed carefully in making the assumptions necessary to develop intermediate solutions. Effectively, assumptions are a form of imputation of missing data: the missing configurations do not have an outcome value, so they are assigned one on the basis of counterfactual reasoning. The counterfactual reasoning takes the form of assuming that the remainder row either would or would not obtain the outcome if cases were to be observed. Frequently, this reasoning takes the form of directional expectations concerning individual conditions, though of course the difficulty here is that the direction is assumed to be the same regardless of context, i.e. regardless of the values of the other conditions. This may or may not be plausible, depending on the particular research situation, but especially given QCA's stress on causal complexity, it has to be well justified.

In this section, I will draw on and expand the analysis published in Glaesser and Cooper (2011) to illustrate the consequences of various assumptions in the face of limited diversity. One of the outcomes analysed in that paper was achieving a higher school qualification than

⁴ Clearly, it can be difficult to know whether more cases could in principle exist or not. However, sometimes we do know that more cases could exist in principle, or even that they do in fact exist. If that is the case, researchers must not make assumptions which contradict this knowledge. Schneider and Wagemann (2012), in their proposed Enhanced Standard Analysis (ESA), use an existing truth table to identify conditions which they deem “empirically necessary”. However, this designation of conditions as empirically necessary only holds if the remainder rows can be safely assumed never to exist. But Schneider and Wagemann then proceed to incorporate these so-called empirically necessary conditions in solutions applicable to the remainder rows. Thus, they effectively assume that these rows could have cases, i.e. that they could exist. But to identify these empirically necessary conditions, it had to be assumed that the remainder rows could *not* exist (for an extended discussion and demonstration of why ESA is flawed, see Cooper and Glaesser 2016). Thus, the assumptions made are *not* consistent within one analysis, and the “empirically necessary conditions” are not in fact empirically necessary, since to identify them, an assumption has to be made which goes beyond the evidence.

Table 2 Truth table with outcome “moving up from Hauptschule”

row no.	MALE	HSREC	SC1P	RS1P	OUT	n	consistency
1	0	0	1	1	1	9	0.889
2	0	0	0	1	1	15	0.867
3	1	0	1	1	1	15	0.800
4	1	0	0	1	0	17	0.765
5	1	0	1	0	0	3	0.667
6	0	0	0	0	0	8	0.500
7	1	1	1	0	0	2	0.500
8	1	0	0	0	0	7	0.429
9	0	1	0	0	0	15	0.400
10	1	1	1	1	0	8	0.375
11	1	1	0	0	0	17	0.294
12	0	1	0	1	0	11	0.273
13	1	1	0	1	0	13	0.231
14	0	1	1	1	0	6	0.167
15	0	1	1	0	0	2	0.000
16	0	0	1	0	0	1	0.000

Hauptschulabschluss if Hauptschule was the type of secondary school attended initially. In the German tripartite system, the Hauptschule is the most basic type of school. Data employed were from the German Socio-Economic Panel (SOEP), with the number of cases for this part of the analysis $n=149$. Clearly, this is a scenario where a great number of cases other than those in the analysis exist and where generalisation beyond the SOEP sample is desirable.

The four conditions used are MALE (respondent is male=1, respondent is female=0), SC1P (respondent has at least one parent in the salariat social class=1, no parent in the salariat=0), RS1P (respondent has at least one parent with at least Realschulabschluss, the qualification obtained at the intermediate type of secondary school=1, both parents have Hauptschulabschluss=0), and HSREC (respondent received a recommendation at the end of primary school for Hauptschule=1, recommendation was for a higher type of secondary school=0). Thus, with four binary crisp conditions, we have 16 truth table rows. In the analysis reported in Glaesser and Cooper (2011), there is no limited diversity, though some rows have very low n , as can be seen in the truth table reproduced here as Table 2. Rows with n smaller than 10 are shaded in grey.

Glaesser and Cooper (2011) obtained three Boolean solutions with different cut-off points for consistency. Since they did not employ a threshold for the number of cases, instead allowing for all the truth table rows to enter the minimisation process, they did not have to choose between complex, intermediate and parsimonious solution types. Here, I focus on the solution they obtained from employing a 0.8 cut-off point. Since the parameters of fit are not relevant for the discussion which follows, I will omit them and merely present the models. The result of the Boolean minimisation, with all rows included and a threshold for consistency of 0.8, was $hsrec*SC1P*RS1P+male*hsrec*RS1P \rightarrow HSUP$. Thus, there are two pathways to the outcome, one involving a recommendation for a higher school type than Hauptschule combined with high parental educational and high parental class status, the other, for females, a higher recommendation combined with high parental education.

Table 3 Original, complex, intermediate and parsimonious solutions

Original solution	hsrec*SC1P*RS1P+ male*hsrec*RS1P=>HSUP
Complex solution	male*hsrec*sc1p*RS1P+MALE*hsrec*SC1P*RS1P=>HSUP
Intermediate solution	hsrec*SC1P*RS1P + male*hsrec*RS1P => HSUP
Parsimonious solution	SC1P + male*hsrec=>HSUP

We may well assume – as Glaeser and Cooper (2011) did – that it is not necessary to exclude any rows from the minimisation process, regardless of n. However, because of the possibility of measurement error, it may be considered safer not to rely on rows with low n and therefore to exclude such rows. Accordingly, I will now repeat the analysis, but only include rows containing at least 10 cases. The rows thus treated as remainder rows have been shaded in grey in Table 2. The original model may serve as a comparison point for the resulting models. To obtain the parsimonious solution⁵, no assumptions concerning the data are required, though it does rest on a particular causal model and on having parsimony as the aim of the analysis. For the complex solution, we have to assume that *none* of the empty rows (shaded in grey) would have the outcome. This may be plausible for rows near the bottom of the truth table, but less so with row 1 in particular, since it combines a favourable parental background with a higher recommendation (and, in fact, the empirical evidence suggests that cases with this configuration of conditions would indeed obtain the outcome, assuming that nine cases is a large enough evidential basis). For the intermediate solution, I initially make the following assumptions: both high parental education and high social class can be expected to be associated with a positive outcome, and so can a higher recommendation. I make no assumptions regarding gender. I do not expect these assumptions to be particularly controversial, and, indeed, they are borne out by the original model. The resulting models are presented in Table 3. For comparison purposes, the table also contains the original solution based on the full truth table.

All three solution types differ, with, as expected, the intermediate solution a superset of the complex and a subset of the parsimonious solution. The complex solution contains a counterintuitive element: the first pathway contains the *absence* of high parental class even though it makes more theoretical sense to assume that this would be associated with a favourable outcome. In fact, this contradicts the original solution for women since it suggests that they should combine the *absence* of high parental social class with the presence of high parental education, while the original solution specifies the *presence* of high parental social class combined with high parental education for both men and women. In addition,

⁵ I used Ragin’s software fsqca to obtain the solutions reported in this section (Ragin and Davey 2019). Given that fsqca uses the Quine-McCluskey algorithm which relies on assumptions regarding the remainder rows to obtain the parsimonious solution, I have repeated the analysis using Thiem’s R software QCApro which employs a different algorithm (Thiem 2018). The results were the same for both the parsimonious and the complex solutions. Given that Thiem believes that only the parsimonious solution produces “correct” results (Thiem 2019), the intermediate solution is not implemented in his software (the complex solution is retained for “method evaluation”, QCApro manual p.27). A comparison using the cna package (<https://cran.r-project.org/web/packages/cna/index.html>) did not produce the parsimonious solution, yielding hsrec <-> HSUP instead with thresholds set at con=0.8, cov=0.429 (the latter is the coverage found in the original Glaeser and Cooper paper). Identifying possible reasons for the discrepancy is beyond the scope of the present paper; I can only speculate that among them is the fact that the SOEP data are noisy and that cna handles thresholds differently from fsqca and QCApro (see footnote 3).

Table 4 Intermediate solutions under various assumptions

	Assumption (directional expectation)	Intermediate solution
1	MALE	male*hsrec*sc1p*RS1P+MALE*hsrec*SC1P*RS1P=>HSUP
2	sc1p rs1p	male*hsrec*sc1p + MALE*hsrec*SC1P=>HSUP
3	HSREC	MALE*SC1P*RS1P + male*hsrec*sc1p*RS1P=>HSUP
4	MALE sc1p rs1p HSREC	male*hsrec*sc1p*RS1P + MALE*hsrec*SC1P*RS1P=>HSUP

the complex solution does not achieve any reduction; unlike the other two solutions, it merely reproduces the two truth table rows which entered the Boolean solution process. The parsimonious solution is a superset of the original solution. The intermediate solution is identical with the original solution, which suggests that the counterfactual assumptions made were sensible.

However, the point of undertaking an empirical analysis is to gain new insights concerning something for which empirical evidence so far has been lacking. If there are configurations without cases, then clearly empirical evidence for these particular cases is still lacking. Thus, the counterfactual assumptions remain just assumptions, even if plausible and well-grounded in existing theory and/or case knowledge. Therefore, they may turn out to be wrong, either because the existing knowledge was not secure enough or because conditions behave differently in isolation compared to in combination with others. The assumptions I made here regarding the SOEP data were well founded both based on existing theory and in fact, based on the evidence from the full truth table, assuming that we allow this evidence despite the low n in some rows. But for the sake of argument, I will now demonstrate how the intermediate solution changes if other assumptions are made. With some imagination, it would be perfectly possible to supply the reasoning behind the assumptions. No combinations are logically impossible, so all assumptions are permissible in principle.

I do not wish to comment on the substantive merit of these assumptions. Clearly, this is questionable in most cases, and I include some precisely because they were unlikely, so as to illustrate the range of possible intermediate solutions. My aim was instead to point out how strongly the models produced by the intermediate solution depend on the assumptions made.

All four solutions are of course subsets of the parsimonious solution. Solutions 1 and 4 in Table 4 are identical with the complex solution. They and solutions 2 and 3 all include a mixture of plausible and less plausible components. But the point here is that we only happen to know which ones are plausible and which ones less so because, in this particular example, we happen to have fairly secure theoretical knowledge as well as empirical evidence from the original analysis. If we didn't, we would have no way of knowing which of our assumptions are plausible and sensible and which ones would result in models not compatible with the social processes under study. It is worth repeating here that the complex solution also relies on an assumption, which is that the empty rows, if cases were to be found to populate them, would not have the outcome. In the same way as the counterfactual assumptions on which the intermediate solutions are based, this can be more or less plausible, depending on the specific situation.

Including implausible assumptions increases the range of models. Especially in a research situation with little theoretical prior knowledge and little case knowledge, this may be useful so that it is possible to get a sense of the likely range within which the underlying causal model may be found. Making the assumptions explicit is an important step in the research process since knowing about them is crucial in assessing the plausibility of the resulting models. Kogut and Ragin (2006), in working through their examples, take such an approach: they make their assumptions explicit, work through the consequences, and use this to contribute to theory development. Limited diversity, in their paper, is a starting point for in-depth analysis of theory and data. In their own way, Schneider and Wagemann (2012) offer another example of such an approach in that they make explicit their assumptions and work through the consequences, though their goal in introducing ESA (Enhanced Standard Analysis) is to offer one definitive approach to dealing with limited diversity which, given the uncertain nature of all assumptions, may close down possible avenues for theory development rather than open them up.

4 Conclusions

The examples I used in this paper were intended to be merely illustrative rather than exhaustive or representative. That is, I have not presented a complete account of all the possible constellations in which limited diversity might arise, nor have I given an exhaustive overview of the possible assumptions needed in all these constellations. My aim was simply to discuss some of the differing contexts in which limited diversity might arise and stress that this always involves choices and assumptions which impact the results. Accordingly, I suggest researchers dealing with limited diversity make sure they are aware of the need for making choices and assumptions and that they make these explicit, as well as discussing the consequences arising from their choice.

One way of thinking about these issues is to work out which type of mistake we are more willing to accept than others, or which dangers might arise from making a choice which later turns out to have been wrong given additional data. The parsimonious solution, while being the only solution type not to rely on any assumptions (if the table of data is taken as given), can be less precise than the other solution types and it can cover cases which, once relevant additional data have been found, turn out ultimately not to be part of the true model. Thus, the mistake here is to be overly inclusive. Conversely, by choosing the complex solution, the researcher is at risk of being overly cautious and mistakenly excluding configurations which actually do turn out to result in the outcome, once the relevant data have been found. If that happens, then it is because of the erroneous assumption that any remainder rows would not have the outcome. The obvious danger with the intermediate solution is that some or all of the counterfactual assumptions may be mistaken. The result here would be a partially incorrect model which relies on a mixture of empirical evidence and counterfactual assumptions, without it being immediately obvious which part is empirically based and which part is reliant on assumptions. This matters regardless of whether a model based on an intermediate solution is correct or not, something a researcher cannot be certain of anyway, since not all the relevant cases for ascertaining this exist in a situation of limited diversity. This is not to say that such models should not be obtained. On the contrary, such a model may well be better than one based on either the parsimonious or the complex solution

because it is more plausible, more informed by theoretical and case knowledge, but it is not purely empirical (see also Haesebrouck 2022, for a discussion of the relationship between empirical evidence and the intermediate solution).

Apart from assumptions concerning the outcomes of remainder rows, assumptions are made concerning whether more cases could exist or not. The danger with assuming more cases could exist when this is not the case is that an intermediate solution may be obtained which relies on assumptions concerning cases which do not and could not exist. The converse, assuming that no more cases could exist when they could, is that an intermediate solution would then not be obtained even though this might be appropriate.

The upshot of all this is that it is not obvious to me that any one solution type is superior, certainly not to the degree that the others always have to be ruled out. They rely on different kinds of assumptions and models of causation. Thus, depending on the research situation, it may be helpful to analyse different scenarios, including one(s) where we assume that our initial assumptions are wrong (as in Sect. 3.2). Ideally, researchers are able to take steps to reduce or eliminate limited diversity since none of the existing solution types are without problems. However, since this is not always possible, it is best to be aware of the consequences of choices, and my aim was to contribute to an improved understanding amongst researchers of the range of possible consequences.

Acknowledgements My thanks go to Barry Cooper for his helpful comments on this paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Statements and Declarations The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

The author has no relevant financial or non-financial interests to disclose.

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

- Baumgartner, M.: Parsimony and causality. *Qual. Quantity*. 49(2), 839–856 (2015)
- Baumgartner, M., Thiem, A.: Often Trusted but Never (Properly) Tested: Evaluating Qualitative Comparative Analysis. *Sociol. Methods Res.* 49(2), 279–311 (2020). <https://journals.sagepub.com/doi/full/10.1177/0049124117701487>
- Baumgartner, M.: Qualitative Comparative Analysis and robust sufficiency. *Quality & Quantity*. (2021). <https://link.springer.com/article/10.1007/s11135-021-01157-z>
- Cooper, B., Glaesser, J.: Qualitative Comparative Analysis, Necessary Conditions, and Limited Diversity: Some Problematic Consequences of Schneider and Wagemann's Enhanced Standard Analysis. *Field Methods*. 28(3), 300–315 (2016). <https://journals.sagepub.com/doi/https://doi.org/10.1177/1525822X15598974>

- Duşa, A.: Critical Tension: Sufficiency and Parsimony in QCA. *Sociological Methods & Research*. (2019). <https://journals.sagepub.com/doi/10.1177/0049124119882456>
- Glaesser, J., Cooper, B.: Selectivity and Flexibility in the German Secondary School System: A Configurational Analysis of Recent Data from the German Socio-Economic Panel. *Eur. Sociol. Rev.* 27(5), 570–585 (2011). <https://doi.org/10.1093/esr/jcq026>
- Haesebrouck, T., Thomann, E.: Introduction: Causation, inferences, and solution types in configurational comparative methods. *Qual. Quant.* (2021). <https://doi.org/10.1007/s11135-021-01209-4>
- Haesebrouck, T.: Relevant, Irrelevant, or Ambiguous? Toward a New Interpretation of QCA's Solution Types. *Sociological Methods & Research*. (2022). <https://journals.sagepub.com/doi/full/10.1177/004912412111036153>
- Kogut, B., Ragin, C.C.: Exploring Complexity when Diversity is limited: Institutional Complementarity in Theories of Rule of Law and National Systems Revisited. *Eur. Manage. Rev.* 3, 44–59 (2006). <https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1057/palgrave.emr.1500048>
- Ragin, C.C.: *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies*. University of California Press, Berkeley, Los Angeles, London (1987)
- Ragin, C.C.: *Fuzzy-Set Social Science*. University of Chicago Press, Chicago and London (2000)
- Ragin, C.C.: *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press, Chicago (2008)
- Ragin, C.C., Davey, S.: *fsqca software, version 3.1*. Retrieved from (2019). <http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml>
- Ragin, C.C., Sonnett, J.: Between Complexity and Parsimony: Limited Diversity, Counterfactual Cases, and Comparative Analysis. In: Kropp, S., Minkenberg, M. (eds.) *Vergleichen in der Politikwissenschaft*, pp. 180–197. VS Verlag für Sozialwissenschaften, Wiesbaden (2005)
- Schneider, C.Q., Wagemann, C.: *Set-theoretic methods for the social sciences. A guide to Qualitative Comparative Analysis*. Cambridge University Press, Cambridge (2012)
- Swiatczak, M.D.: Different algorithms, different models. *Quality & Quantity*. (2021). <https://link.springer.com/article/10.1007/s11135-021-01193-9>
- Thiem, A.: *QCApro: Advanced Functionality for Performing and Evaluating Qualitative Comparative Analysis. R Package Version 1.1-2*. URL: (2018). <http://www.alrik-thiem.net/software>. <http://cran.r-project.org/package=QCApro>
- Thiem, A.: Beyond the Facts: Limited Empirical Diversity and Causal Inference in Qualitative Comparative Analysis. *Sociological Methods & Research*. (2019). <https://journals.sagepub.com/doi/10.1177/0049124119882463>
- Thomann, E., Maggetti, M.: Designing Research With Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools. *Sociol. Methods Res.* 49(2), 356–386 (2020). <https://journals.sagepub.com/doi/full/10.1177/0049124117729700>

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