Chapter 7

Set-Theoretic Analysis

Mello, Patrick A. (2021) *Qualitative Comparative Analysis: An Introduction to Research Design and Application,* Washington, DC: Georgetown University Press, Chapter 7.

Necessary conditions pose traps to the researcher because of their deceptive simplicity.

Gary Goertz and Harvey Starr [1]

A truth table [...] is a direct examination of the kinds of cases that exist in a given set of data.

Charles C. Ragin [2]

Earlier chapters introduced *set theory* as the foundation on which QCA rests, showed how raw data can be *calibrated* into crisp and fuzzy sets, and discussed *measures of fit* for the assessment of set-theoretic relationships. With these essentials in place, this chapter turns to the set-theoretic analysis of empirical data.

Set-theoretic analysis comprises testing individual conditions and combinations of conditions for their necessity and/or sufficiency. These tests are done separately, beginning with the analysis of necessary conditions before moving to the truth table procedure, which is directed at testing for sufficient configurations and subsequent Boolean minimization. This is the analytical core of QCA.

Analyzing Necessary Conditions

It is considered *good practice* to analyze necessary conditions in a separate step before engaging in the truth table procedure (Ragin 2000, 204; Schneider and Wagemann 2010, 404). Knowing about the existence of a necessary condition helps to make informed choices during the minimization of the truth table and the treatment of logical remainder rows, which we will cover towards the end of this chapter.[3] Moreover, necessary conditions are not directly observable from the analysis of sufficient conditions, especially if the set-theoretic relationship of necessity is imperfect. This may happen, for instance, when there is at least one case that shows the outcome despite the absence of the condition. In such a situation, one might overlook a necessary condition if the focus rests solely on the truth table procedure and its minimization.

As discussed in previous chapters, having a necessary condition means that whenever the outcome occurs, the condition is also present. In formal terms, the outcome is thus a *subset* of the necessary condition and, vice versa, the necessary condition is a *superset* of the outcome. The primary measure of fit for necessary conditions is *consistency*. By convention, conditions have to pass *at least* 0.9 set-theoretic consistency to be considered necessary (Mendel and Ragin 2011, 21; Schneider and Wagemann 2012, 143).

Besides the formal benchmark of set-theoretic consistency, one also has to consider *coverage*, because conditions may be formally necessary but empirically less meaningful or even irrelevant. This is complemented by the parameter *relevance of necessity*, as an additional metric for necessary conditions (Schneider and Wagemann 2012, see the discussion in Chapter 6). Together, the three parameters *consistency*, *coverage*, and *relevance* should be considered when testing for necessity. For conditions that do not pass 0.9 consistency, the other measures are meaningless. But once the consistency threshold has been passed, then one should further consider coverage and relevance to assess the condition under study. As rule of thumb, when the metrics *coverage* or *relevance* are below 0.5, then this suggests that we may be dealing with a *trivial* necessary condition. To make an informed judgment in such a setting, we should always examine the empirical distribution of our cases and their set-theoretic membership scores (as discussed in Chapter 6).

What about *combinations* of conditions? In most situations, you may either have no expectation about a potential necessary condition or, when you do, then you would typically expect a single condition to be necessary on its own. However, in some situations you may also have an expectation about a particular combination of factors, each of which has to be present for the outcome to occur. In logical terms, when we expect several conditions to be *jointly necessary*, then each single element of that conjunction must also be necessary. Hence, a test for individual conditions should suffice to show whether or not the expected conditions are truly necessary for the outcome.

Yet occasionally a theory may propose the *substantive equivalence* of two or more conditions – where the presence of *either of them* may be necessary for the outcome. For example, suppose we are interested in the relationship between party politics and the perceived legitimacy of the European Union. Based on our theory, we may say that in order to increase the EU's perceived legitimacy political parties either need to organize and campaign on the European level (as opposed to electoral campaigns based on national boundaries), or that parties need to agree upon common candidates for leadership positions in the EU before the elections. Both of these would be expected to increase the perceived legitimacy of the EU in the eyes of the public, but *at least one of them* has to be present for the outcome.

The software allows the specification and testing of such *disjunctional expectations* for necessary conditions. However, a note of caution is warranted: while it is possible to search for any kind of necessary combination, I advise against "fishing" for necessary disjunctions when there is no prior expectation about a set-theoretic relationship. We may discover some complex combinations of conditions (linked by a Boolean OR) to be necessary, but without prior theory it will be challenging to interpret such findings and to fill them with meaning. This resonates with advice not to apply the method "mechanically", in a data-driven way, without due emphasis on existing knowledge and substantive interpretation (Ragin 1987, 120; Schneider and Wagemann 2010, 410).

To illustrate the procedure of testing individual conditions for necessity, let us look at an example. Table 7.1 shows fuzzy-set values for five cases across three conditions (A, B, C), and the outcome Y. The right-hand side of the table displays the results for the calculation of set-theoretic consistency, coverage, and relevance of necessity (RoN), based on the formulae introduced in Chapter 6. We can see that all three conditions pass the formal threshold of 0.9 consistency. In fact, all of them have *perfectly consistent* set relations of necessity, because the set-membership values for each condition and case are equal to or exceed those for the outcome. This means that A, B, and C are all *supersets* of the outcome Y.

However, although all three conditions pass the formal benchmark for the consistency of necessary conditions, their coverage and relevance scores indicate that some of them are more important than others. For condition A, both coverage and RoN are close to 1 (0.8 and 0.83), which means that, based on these metrics and without considering the substantive background, A can be considered a relevant necessary condition. Condition B has a fairly low coverage (0.54) and even lower RoN (0.43). This suggests that condition B is clearly a *less relevant* necessary condition than condition A. Whether condition B should further be considered a trivial necessary condition is a matter of substantive interpretation for which we would have to take into account a study's theory and research context. Finally, with condition C the pattern is more pronounced. C has a slightly lower coverage than B (0.43), but the RoN value drops close to zero (0.1). Looking at the values for condition C across the five cases, we can see that C is *always present* (all cases hold membership values above 0.5). Hence, there is not much gained from saying that "C is a necessary condition for Y" because C is a constant in this setting. Hence, given this data, C should be considered a *trivial* necessary condition.

Table 7.1 Necessary Conditions: Consistency, Coverage, Relevance

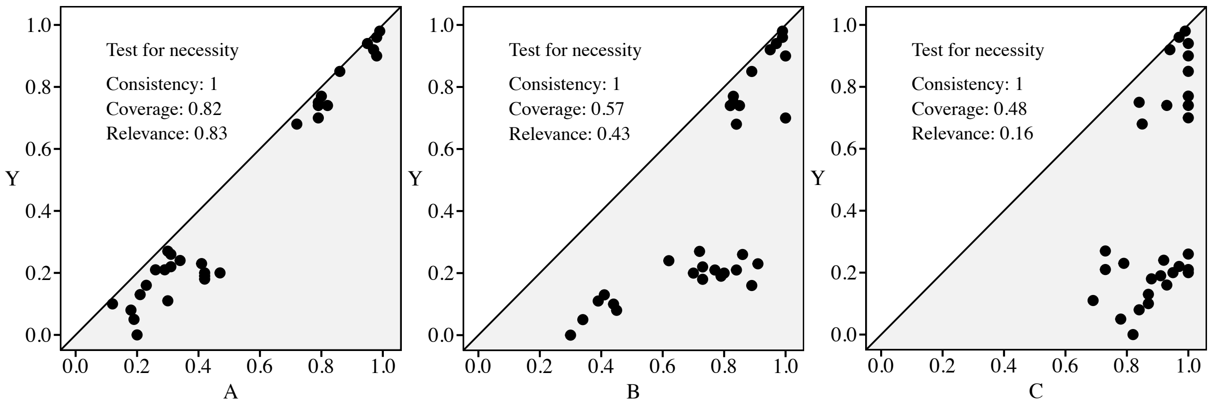
	Conditions			Outcome	Test for necessity			
	A	B	C	Y		Consistency	Coverage	Relevance
Case 1	0.9	1	1	0.9	A	1	0.80	0.83
Case 2	0.8	1	1	0.7	B	1	0.54	0.43
Case 3	0.3	0.7	1	0.2	C	1	0.43	0.10
Case 4	0.4	0.8	0.9	0.2				
Case 5	0.1	0.2	0.8	0				

Figure 7.1 illustrates these relationships of necessity for a larger number of cases (30 cases across conditions A, B, C). We can see that all three conditions fit the measure of set-theoretic consistency, because all of them are located on or below the main diagonal. However, as can be seen from the three XY plots, there are considerable differences in coverage and relevance of necessity. While condition A in the first plot should be considered a relevant necessary condition, the pattern in the second plot indicates that condition B is a less relevant necessary condition. Finally, condition C clearly is a trivial necessary condition, as indicated by a low coverage, and particularly by its relevance of necessity, which is close to zero.

What we can also glean from the plots is how certain values in the metrics correspond to specific patterns in the data. For relevant necessary conditions, the cases are located closer to the diagonal line, whereas irrelevance increases the further cases are lined up on the right-hand axis. This becomes evident with condition C, which approximates a constant where the cases populate a small range of values for the condition, between 0.7 and 1, whereas they show the full range of values for the outcome. This is evidence against a meaningful relationship between the condition and the outcome. Hence, condition C should be considered a *trivial necessary condition.*

To sum up, the analysis of necessity should *precede* the truth table analysis. Each condition and its negation should be tested for their necessity vis-à-vis the outcome and the non-outcome. For studies that focus on explaining the outcome, it suffices to document tests for necessity for the outcome. Yet, due to causal asymmetry it is recommended to also conduct and report separate tests for the non-outcome, with all conditions and their negation. As the primary measure of fit *consistency* determines whether the other metrics are examined (since these are not meaningful otherwise). Conditions with *at least* 0.9 consistency can be considered necessary conditions *if* they also have high enough coverage and relevance of necessity scores. Else they may be considered trivial and should not be treated as necessary conditions. How *high* do coverage and relevance of necessity have to be? In the absence of firm benchmarks for coverage and relevance, a rule of thumb is that coverage and relevance scores *below* 0.5 indicate that the respective necessary condition may be trivial.[4] Against that backdrop, the illustrative data patterns shown in Table 7.1 and Figure 7.1 are meant to provide yardsticks for the comparison of empirical relationships. But as with all metrics, the numerical indicators should never replace *substantive interpretation* and *case knowledge* when trying to "make sense" of one's analytical findings.

Figure 7.1 Necessary Conditions: Consistency, Coverage, and Relevance



Truth Table Construction

The core of QCA is the truth table analysis. As the method is grounded in a *combinatorial logic*, the truth table represents the number of combinations that are logically possible with the selected number of conditions. Each row of the truth table stands for a specific combination of conditions (*configuration*) and the number of rows equals the overall number of possible configurations. The formula for this is 2^k, where k is the number of conditions included in a study. Hence, when the analysis comprises three conditions, then there are eight rows of configurations, four conditions result in 16 rows, and five conditions yield 32 rows, and so forth. We can see that the truth table *grows exponentially* with every additional condition. This distinguishes truth tables from a conventional spreadsheet where each row represents a single observation or case, and where the size of the data matrix is circumscribed by the number of observations and variables.

The truth table shows possible configurations, but it also provides information about the empirical distribution of cases, and their connection to the outcome. Therefore, each row of the truth table is also a *statement of sufficiency*. Some rows may consistently lead toward the outcome, where each case with the respective combination shows the phenomenon of interest, whereas other rows might have no cases that show the outcome. There can also be *contradictory rows* in the truth table, where some cases have a positive outcome and others a negative outcome (Rihoux and De Meur 2009; Rubinson 2013; Yamasaki and Rihoux 2009).

We can illustrate the construction of a truth table with a simple example. Table 7.2 shows a truth table for crisp set data on 20 cases, the conditions A, B, C, and the outcome Y. Based on previous research, we expect the three conditions to be sufficient for the outcome, either individually or in combination. The table shows all logically possible combinations of the three conditions, where each row reflects one specific combination. In total, there are eight rows, as 2^3 equals 8. For instance, row 6 refers to the absence of condition A, presence of condition B, and absence of condition C. There are two empirical cases that share this combination of conditions and both cases do not show the outcome Y. We can also see that one combination (row 7) is not filled with an empirical case. This means that there is no case that holds membership above 0.5 in this configuration. Hence, we cannot say whether this row is connected to the outcome or the non-outcome (indicated by the question mark). This is the issue of *limited diversity* that we will return to when discussing solution terms. Finally, we can see that the combinations in the first three rows of the truth table are associated with the outcome. These will be the rows to focus on for an explanation of the phenomenon of interest.

The truth table provides valuable information about the empirical distribution of cases –

the "kinds of cases" that Ragin refers to in this chapter's introductory quote. The truth table also helps to think systematically about theoretical expectations. For instance, what about the configuration in row 7? If there were an empirical case with membership above 0.5 in this configuration, would this case show the outcome? And why is it that row 4 does not have a positive outcome, although two of the three conditions we expected to lead to the outcome are present?

The truth table helps to analytically explore the data and to formulate such questions, which is particularly useful when there are even more conditions and possible configurations to consider. During the early phases of designing a QCA study, a preliminary truth table can be used to guide the selection of conditions because it shows whether or not the included conditions are *difference-makers* that adequately distinguish the observed cases. If many cases cluster in just a few rows of the truth table and the remaining rows are empty, then one should explore whether some conditions should be re-conceptualized, re-calibrated, or replaced by conditions previously omitted. Such changes in research design can help to differentiate between clustered cases and result in a more evenly spread distribution across the truth table rows.[5] Likewise, if each case has a "row of its own", then this indicates that the research design may be overly differentiated, a situation that can be improved, for instance, by eliminating certain conditions or by merging some of them into macro-level conditions (see Chapter 3).

Table 7.2 A Simple Truth Table

Row	Conditions			Outcome	Cases
#	A	B	C	Y	*N*
1	1	1	1	1	4
2	1	1	0	1	5
3	1	0	1	1	2
4	0	1	1	0	1
5	1	0	0	0	3
6	0	1	0	0	2
7	0	0	1	?	0
8	0	0	0	0	3

Truth Table Analysis

The truth table is in itself a valuable tool for descriptive data analysis. However, since each row in the truth table is a statement about sufficiency, we can use Boolean algebra to minimize the truth table to derive solution terms for an outcome, which is what QCA is all about. This section summarizes and illustrates the Boolean minimization procedure that was developed by Ragin (1987, 93-102; 2000; 2008) and consolidated by Schneider and Wagemann (2012, 104-15; 2013), among others (see also Duşa 2019; Oana et al. 2021). The final sections of this chapter also look into some refinements of what is known as the *standard analysis*.

The truth table analysis is done via software, but the procedures rest on Boolean algebra and as such they can be reproduced by hand. The technical steps are described in the R Manual in Appendix. Let us take an example from a published study, simplified for our purposes, to illustrate the analysis of the truth table. In his article on the politics of religion, Luis Felipe Mantilla (2012) explores the conditions under which the Catholic church promoted democratization in South America.[6] The study uses eight countries as cases and the explanatory conditions *resources*, *opportunity*, and *cultural framing* to explain the outcome *democratization support*, as shown in Table 7.3.

Table 7.3 Data Frame: Politics of Religion

Country	Resources (R)	Opportunity (O)	Framing (F)	Outcome (Y)
Brazil	1	1	1	1
Chile	1	1	1	1
Peru	1	1	0	1
Ecuador	0	1	0	0
Bolivia	0	0	1	0
Uruguay	0	1	1	0
Argentina	1	1	1	1
Paraguay	0	0	1	0

Data source: Mantilla (2012).

Based on this crisp-set data, we can construct a truth table. Table 7.4 shows how the empirical cases distribute across the logically possible configurations. We can see that the eight cases fill five rows, two of which lead consistently toward the outcome (the top two rows). There are also three *empty rows* at the bottom of the table, indicated by a question mark. These are combinations of conditions without empirical cases in the data and thus, we cannot say whether or not the respective configurations consistently lead to the outcome (these rows are also termed *logical remainders*). Yet, one advantage of QCA is that it allows for counterfactual reasoning through a systematic treatment of these logical remainder rows. This is the issue of *limited diversity* that we will return to when discussing solution types.

Table 7.4 Truth Table: Politics of Religion

Row	Resources (R)	Opportunity (O)	Framing (F)	Outcome (Y)	*N*	Country
1	1	1	1	1	3	Argentina, Brazil, Chile
2	1	1	0	1	1	Peru
3	0	0	1	0	2	Bolivia, Paraguay
4	0	1	1	0	1	Uruguay
5	0	1	0	0	1	Ecuador
6	1	0	1	?	0	–
7	1	0	0	?	0	–
8	0	0	0	?	0	–

How to proceed with the truth table *analysis*? In a manual application, the first step would be writing down the configurations that are linked to the outcome. In our example, these are the first two rows of Table 7.4, both of which show the outcome. The cases included in these rows are Argentina, Brazil, and Chile for row 1, and Peru for row 2.

We write down the combinations as individual conditions linked by a Boolean AND (·). For convenience, we may also omit this operator and simply write the letters that represent the conditions next to each other, as in "ROF" to indicate the combination in the first line. The combinations themselves are linked by a Boolean OR (+), to indicate the existence of multiple paths, where each is sufficient for the outcome. Hence, the formal notation for the conditions *resources* (R), *opportunity* (O), *framing* (F), and the outcome *democratization support* (Y) is:

(1) $R{\cdot}O{\cdot}F + R{\cdot}O{\cdot}{\sim}F \rightarrow Y$

This reads as "resources *and* opportunity *and* framing *or* resources *and* opportunity *and not*-framing are sufficient for democratization support". To simplify this expression, we can apply the rule of *Boolean minimization*:

> If two Boolean expressions differ in only one causal condition yet produce the same outcome, then the causal condition that distinguishes the two expressions can be considered irrelevant and can be removed to create a simpler, combined expression (Ragin 1987, 93).[7]

Based on this rule, we can delete F and ~F from expression (1), because both combinations lead toward the outcome (with and without the condition *framing*). In the next step, we can eliminate one of the duplicate expressions of R·O, which leaves us with:

(2) $R{\cdot}O \rightarrow Y$

This reads as "resources *and* opportunity are sufficient for democratization support". In other words, when both R and O are present, then this leads to Y. We can see that this solution matches the information contained in the truth table (Table 7.4). There are only two possible combinations that entail the expression R·O (rows 1 & 2) and both of these show the outcome.

This is a simple example, but it helps to illustrate the way Boolean minimization works. Each truth table row that is associated with the outcome yields a *primitive expression*. This is the Boolean notation of the combination contained in the respective row. If there are two or more primitive expressions (truth table rows associated with the outcome), then the minimization rule can be applied to derive a simpler expression, which consists of so-called *prime implicants*.

In another step, this can be further simplified by looking whether some of the prime implicants are *redundant* because primitive expressions may be covered by several of the prime implicants.

To illustrate this point, let us imagine a truth table that yields four combinations of the conditions A, B, C that are sufficient for the outcome Y:

$$(3) \qquad A{\cdot}B{\cdot}C + A{\cdot}B{\cdot}{\sim}C + {\sim}A{\cdot}B{\cdot}C + {\sim}A{\cdot}{\sim}B{\cdot}C \longrightarrow Y$$

Again, we can apply the rule of Boolean minimization to derive a shorter statement. To do so, each primitive expression can be compared to one or several of the other expressions until no further minimization is possible. In the following minimization, the brackets indicate which two expressions from line (4) are used to create the simpler statement in line (5). Note that the procedure shown here does not exhaust all possible comparisons, but these will not yield other solutions. For the minimization, each expression must be compared to at least one other expression or carried over to the next row. We can see that the first expression A·B·C is compared to both the second expression A·B·~C and to the third expression ~A·B·C:

$$(4) \qquad A{\cdot}B{\cdot}C + A{\cdot}B{\cdot}{\sim}C + {\sim}A{\cdot}B{\cdot}C + {\sim}A{\cdot}{\sim}B{\cdot}C \longrightarrow Y$$

$$(5) \qquad A{\cdot}B \quad + \quad B{\cdot}C \quad + \quad {\sim}A{\cdot}C \quad \longrightarrow Y$$

This first step in the minimization procedure eliminates expressions that are deemed *superfluous* for the solution because their presence *and* absence both lead to the outcome, given the presence of the other conditions. This applies, for instance, to the conditions C and ~C in the first two expressions in line (4). They are part of the combinations A·B·C and A·B·~C. Hence, when and A and B are both present, then it does not matter whether they combine with C or ~C, because both combinations lead to the outcome. Hence this can be minimized to A·B, as stated in line (5). The other two minimizations follow the same logic.

For the second step of the Boolean minimization procedure a *prime implicant chart* is constructed (Ragin 1987, 97). Again, this process is implemented in the software, but it helps to go through the sequence in order to understand how QCA solution terms are derived. Users should also be familiar with the prime implicant chart because they may need to work with it (for instance in the fs/QCA software, but also in R). The prime implicant chart displays as columns all truth table rows that are sufficient for the outcome (the *primitive expressions*). The results from the first minimization step (the *prime implicants*) are listed as rows. Table 7.5 shows the chart for the hypothetical truth table data used above:

Table 7.5 Prime Implicant Chart

Primitive Expressions					
A·B·C	A·B·~C	~A·B·C	~A·~B·C		
×	×			A·B	Prime Implicants
×		×		B·C	
		×	×	~A·C	

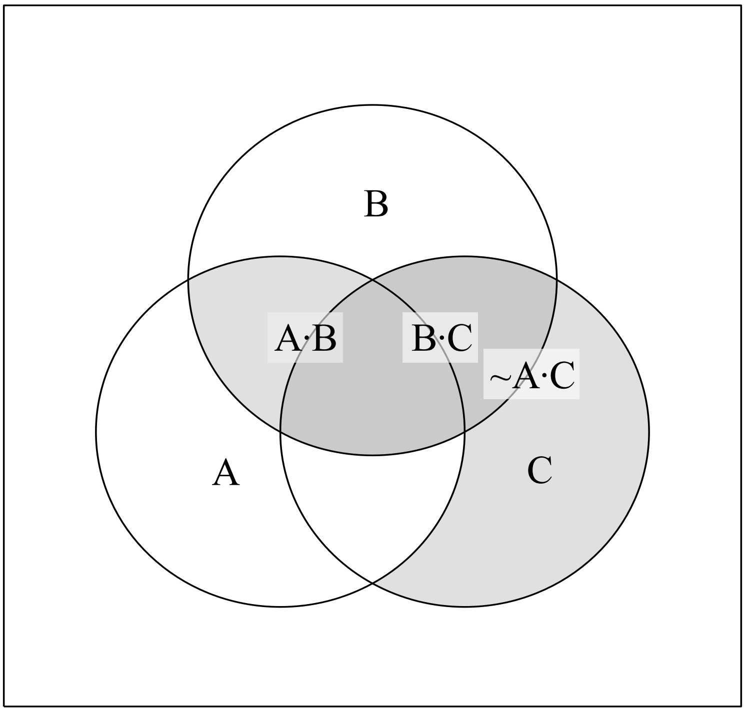
The columns of the prime implicant chart list the configurations (truth table rows) that were linked to the outcome. The crosses in the cells indicate which expression is covered by which prime implicant. This means that the respective prime implicant is a *superset* of the covered primitive expression(s). In our example, we can see that each prime implicant covers two primitive expressions, but the second prime implicant is *redundant* because its primitive expressions are also covered by the other two prime implicants. In short, there is no information gained from keeping the expression B·C in the solution term and thus it can be *deleted* to yield a more parsimonious solution:

(6) A·B + ~A·C ⟶ Y

Why can we erase the expression B·C from the solution? The Venn diagram in Figure 7.2 shows why the shorter statement in line (6) is equivalent to the statement in line (5) above. Quite simply, there is nothing gained from saying B·C, because its gray shaded area in the Venn diagram is *entirely covered* by the other two prime implicants A·B and ~A·C. Hence, in this setting, saying that "A *and* B *or not*-A *and* C are sufficient for Y", *implies* that B·C is also sufficient for Y. B·C is thus a redundant prime implicant.

The prime implicant chart also helps to understand why different solution terms might be derived from the same data. This phenomenon is known as *model ambiguity*, a concept that describes that, depending on the structure of the truth table and the assumptions made, several logically equivalent solution terms may be derived from the same data (Baumgartner and Thiem 2017). Importantly, even when there are multiple solutions, in principle, these are all equally "correct" in the sense that they are the result of an accurate Boolean minimization procedure, based on the empirical information that is entailed in the truth table. However, some solutions may include *redundant elements* and other solutions may entail *untenable simplifying assumptions* about logical remainder rows. These aspects will be examined in the next section (see also Chapter 9).

Figure 7.2 Venn Diagram: Prime Implicants and Redundancy



Solution Terms

The two-stage minimization procedure described in the previous section is incorporated into QCA's software routine which draws upon the *truth table algorithm* to minimize expressions and to derive solution terms. Initially, this was based upon the Quine-McCluskey (QMC) algorithm and the enhanced Quine-McCluskey (eQMC) algorithm, which have been replaced by the Consistency Cubes algorithm (Duşa 2018) as the default minimization method in recent versions of the QCA package (Duşa 2019). From a user's perspective, the differences between these algorithms mainly concerns their performance and reliance upon computer resources, as they reach identical minimization results.[8]

Before the minimization, the researcher has to specify a *consistency threshold* to indicate which rows of the truth table should be treated as positive instances of the outcome and included in the analysis. For these, the outcome column is coded as 1. Rows below the threshold are coded as 0. The standard threshold for this step is 0.75 consistency. Apart for exceptional reasons,[9] researchers should not include rows with a consistency lower than this conventional threshold (Ragin 2008, 144; Rubinson et al. 2019, 4; Schneider and Wagemann 2012, 292). What is important to understand about the consistency threshold is that it merely defines the *lower bound* that determines which rows shall be included in the minimization. Hence, the eventual solution consistency will typically be *higher* than the consistency threshold because there will also be rows of a higher consistency.

Users can further define a *frequency threshold* to indicate how many cases with membership greater than 0.5 must be entailed in a truth table row for it to be included in the minimization procedure. Usually, this is set to 1, which means that a single case suffices, but in large-*N* research settings with high numbers of cases it can be expedient to increase the frequency threshold so that only those rows are included that hold a certain number of cases with membership above 0.5 in the respective configuration. For example, in their large-*N* QCA study on the persistence of electoral autocracies (with 156 cases of elections across 48 autocratic regimes), Carsten Schneider and Seraphine Maerz (2017, 222) select a frequency cutoff of *at least two* empirical cases per row as this helps to guard against the chance of misclassified cases. Notably, the frequency threshold did not substantively alter their QCA results.

Once the consistency and frequency thresholds have been specified, the QCA software is applied to derive three different solution terms: the *conservative solution*, the *parsimonious solution*, and the *intermediate solution*. The differences between these solution types originate from their treatment of logical remainders.

The Conservative Solution

The *conservative solution* has its name because it only works with the empirical rows that are associated with a positive outcome, above the specified consistency and frequency thresholds. As the term *conservative* conveys, this solution does *not* make assumptions about empty rows in the truth table and thus it sidesteps counterfactual reasoning. In fact, the conservative solution treats all logical remainder rows as false (Ragin 2008, 173; Schneider and Wagemann 2012, 162). One consequence of this approach is that it tends to yield lengthier solution terms than the other types. This is so because the conservative solution can use only rows with empirical cases above the specified thresholds for the pairwise Boolean minimization, whereas the other solutions both incorporate logical remainder rows to different extents. This is why the conservative solution is also called the *complex solution* (Ragin 2008, 173). Complexity in this sense refers to the number of conditions and operators that are entailed in a solution term. The more conditions and operators, the greater the complexity of the solution. As Adrian Duşa (2019, 177) notes, "no other solution can be more complex than the conservative one", yet other solutions can be *similarly* complex.[10]

The Parsimonious Solution

The *parsimonious solution* considers all truth table rows with logical remainders and uses those that allow it to derive a less complex solution through pairwise Boolean minimization. Those logical remainders that are used for minimization are called "simplifying assumptions" (Ragin 2000, 305; 2008, 136). These logical remainders are *simplifying* because they enable further minimization through Boolean comparisons. Some logical remainders may have no utility for the minimization, but these would simply not be used by the parsimonious solution. Those logical remainders that are used are further called *assumptions* because using them rests on the counterfactual conjecture that if cases with the respective configuration existed, then these would show the outcome. Since the parsimonious solution is allowed to examine all logical remainders, it can work with the broadest pool of configurations and thus its minimization procedure tends to yield the least complex solution terms. Yet, what the parsimonious solution does *not* consider is the *plausibility* of the simplifying assumptions that were used to derive it. This means that the parsimonious solution may rest on unrealistic assumptions about hypothetical data. Hence, the parsimonious solution should always be scrutinized for its simplifying assumptions (Ragin 2008, 162), which means examining the logical remainder rows that were included in its calculation. We will return to this point in the next section.

The Intermediate Solution

Finally, as its name implies, the *intermediate solution* is positioned between the conservative and the parsimonious solutions because it includes logical remainders, but only those that are deemed sensible. With the intermediate solution, the researcher can decide which logical remainder rows to include in the minimization and which ones should be excluded from the procedure. As Ragin (2009, 118) puts it, the intermediate solution "incorporates only the logical remainders that are consistent with theoretical and substantive knowledge", as defined by the researcher. Here, as a heuristic device, a distinction is made between *easy* and *difficult counterfactuals* (Ragin 2008, 160-67; Ragin and Sonnett 2005). Both of these are simplifying assumptions that are used for the parsimonious solution. Easy counterfactuals are defined as those assumptions that are theoretically meaningful and reasonable in light of case knowledge. A short-cut to the assessment of easy counterfactuals are "directional expectations" about the presence or absence of a condition and its respective relationship to the outcome (Ragin 2003b, 9).[11]

For example, based on theory, we may expect the *presence* of an electoral system that is based on proportional representation to lead to a multiparty system. In principle, logical remainders where such an electoral system is present should thus be considered easy counterfactuals. By contrast, *difficult counterfactuals* are assumptions that conflict with theory and substantive knowledge. For instance, this may be the case when a logical remainder entails the *absence* of conditions whose presence is expected to lead to the outcome. The basic idea behind the intermediate solution is to include easy counterfactuals while excluding difficult counterfactuals from the analysis (Ragin 2008, 174).

How is this done in practice? One way to derive an intermediate solution is by *excluding* specific truth table rows (Schneider and Wagemann 2012, 171). This can be fine when there are few simplifying assumptions but complicated with large truth tables that contain many empty rows.

Another way to do this is through the formulation of *directional expectations* for each condition (Schneider and Wagemann 2012, 168). Based on this information, the algorithm in the software then selects simplifying assumptions in line with the formulated expectations (Duşa 2019).

However, even though the software takes directional expectations into account, this does not mean that all counterfactuals that meet the criteria will be considered in the calculation of the intermediate solution term. The scope for the intermediate solution is *circumscribed* by the other two solution terms. The starting point is the conservative solution, which rests entirely on the empirical cases. The parsimonious solution seeks to arrive at simpler expressions, but it may be based on unrealistic assumptions about certain logical remainder rows. For the intermediate solution, those logical remainders that are deemed *plausible counterfactuals* may be used for the minimization, which should in principle result in a solution term that is situated between the conservative and the parsimonious solution in terms of complexity (see also the next section on counterfactual analysis).[12]

Solution Terms: An Illustration

The relationship between the three different types of solution terms is best illustrated with an example. Table 7.6 shows a truth table for fuzzy-set data across three conditions and 25 cases. Note that the software indexes the truth table rows and that these designators are kept throughout the analysis. Since the truth table is sorted by consistency and frequency, the numerators are not in sequence any longer, but each row retains its *unique identifier* throughout the analysis. This may become important, for instance, when seeking to exclude a certain logical remainder row from the minimization.

From Table 7.6, we can see that three rows are associated with the presence of the outcome. We can summarize this information in Boolean notation:

(7) A·~B·C + ~A·B·C + A·B·~C ⟶ Y

The truth table further shows that there are three logical remainder rows, as in rows without empirical cases. What can we say about these? Given the pattern in the first three rows and the way the conditions were calibrated, we may expect that the joint presence of all three conditions would also lead to the outcome if such cases existed. This applies to the logical remainder in row 8 (A·B·C). which is thus an *easy counterfactual*. But what about row 1 (~A·~B·~C) and row 3 (~A·B·~C)? Both of these could be considered *difficult counterfactuals* because the conditions point in the "wrong" direction. We also know from the rows associated with the outcome that it seems to require at least two of the three conditions to bring about the outcome, so the absence of all of them (row 1) or two of them (row 3) may not suffice for the outcome.

Table 7.6 Example Truth Table

Row	A	B	C	Outcome	*N*	Consistency	PRI
6	1	0	1	1	6	0.95	0.95
4	0	1	1	1	5	0.95	0.94
7	1	1	0	1	5	0.95	0.94
2	0	0	1	0	5	0.37	0.16
5	1	0	0	0	5	0.26	0.14
1	0	0	0	?	0	–	–
3	0	1	0	?	0	–	–
8	1	1	1	?	0	–	–

We can now derive the three solution terms from the truth table. In this example, the *conservative solution* is identical to the top three rows of the truth table because these cannot be further minimized without using logical remainders:

(8) A·~B·C + ~A·B·C + A·B·~C ⟶ Y

Next, we derive the *parsimonious solution*, for which the software considers all logical remainders and uses those that enable it to derive a simpler, more parsimonious solution term. This process yields the following solution:

(9) A·C + B ⟶ Y

This parsimonious solution rests on simplifying assumptions about logical remainder row 3 (~A·B·~C) and row 8 (A·B·C), both of which are treated as if they were sufficient for the outcome. How is this result achieved? The minimization that is entailed in the parsimonious solution can be reproduced manually by adding the two expressions for the logical remainders. I should highlight that what follows is a *stylized example* that simplifies the algorithmic routine to illustrate what is *effectively* going on with the parsimonious solution. First, the order in which the expressions appear is sorted for a more convenient illustration:

(10) $A\cdot{\sim}B\cdot C + A\cdot B\cdot C + A\cdot B\cdot{\sim}C + {\sim}A\cdot B\cdot C + {\sim}A\cdot B\cdot{\sim}C \longrightarrow Y$

(11) $A\cdot C + B\cdot C + B\cdot{\sim}C + {\sim}A\cdot B \longrightarrow Y$

(12) $A\cdot C + B + {\sim}A\cdot B \longrightarrow Y$

(13) $A\cdot C + B \longrightarrow Y$

Through pairwise Boolean comparison, indicated by lines and brackets, we reduce the original term of five expressions (10) to four expressions in statement (11). From here, we can further eliminate ~C and C, which yields the three expressions in statement (12). The final minimization rests on Boolean logic because the expression ~A·B in statement (12) is covered by the single condition B. This can be confirmed with a prime implicant chart or a Venn diagram (see earlier sections in this chapter). Hence the parsimonious solution is the statement in line (13), which cannot be further minimized.

However, we know that this parsimonious solution rests on the inclusion of a difficult counterfactual (row 3). Hence, we create an *intermediate solution*, which allows us to specify which logical remainder rows should be used. Here, we may want to treat row 8 as an *easy counterfactual* because all conditions are present, but we specify that the software should *exclude* row 3 from being used in the solution because this constitutes a difficult counterfactual with the presence of only a single condition. If we pass this criterion on to the software, it returns the following intermediate solution, which rests on the inclusion of a single logical remainder (row 8):

(14) $A\cdot C + B\cdot C + A\cdot B \longrightarrow Y$

Table 7.7 shows the three different solution types and their measures of fit. This confirms that the parsimonious solution is the *superset* of the other solutions and that the conservative solution is a *subset* of the intermediate and parsimonious solutions. This general relationship between the solution types will always apply, the only exception being identical solutions. In this example, we see that the three paths that comprise the conservative solution are fully covered by the two paths of the parsimonious solution (A·~B·C is covered by A· C, whereas ~A·B·C and A·B·~C are both covered by B). The same principle applies for the intermediate solution, which is a *superset* of the conservative and a *subset* of the parsimonious solution.

As for measures of fit, in this example the intermediate and conservative solutions yield nearly identical scores, whereas the parsimonious solution has slightly higher coverage and less consistency than the other solutions. Depending on the empirical data, the differences may be more pronounced. Typically, the parsimonious solution has the highest coverage, whereas the conservative solution tends to have the highest consistency.

The relationship between the solution types is summarized in Figure 7.3, which places the three solutions on a single dimension that runs from *complexity to parsimony*, whereas the Venn diagrams in Figure 7.4 depict the subset/superset relations that govern the solutions.[13] The illustration shows that the parsimonious solution is the *most general* as it covers the largest area in the diagram. Yet, we should note that this parsimonious solution is based on a difficult counterfactual. Hence it should always be scrutinized and compared to the other solutions, as discussed above. The conservative and intermediate solutions are subsets of the parsimonious solution and hence it can be said that they are *more specific*, which also shows in the smaller areas covered in the Venn diagrams.

Which solution type should be given preference? As this section has shown, the solution types are not ordered in a hierarchical fashion but rather characterized by their *interdependence*.

In general, *all three solutions* should be derived and examined. Ideally, the three solutions should also be reported in publications, either in the main text or in online supplements. That said, published studies typically emphasize a single solution type in their substantive interpretation. Most often, this is the *intermediate solution* because it allows researchers to determine the treatment of logical remainders based on their theoretical expectations and substantive knowledge of the research area. This is also why the intermediate solution has been emphasized in the QCA literature (Ragin 2008, 175; Schneider and Wagemann 2012, 279). However, the *conservative* and *parsimonious solutions* are equally valid, under the precondition that the latter is checked for untenable assumptions. After all, the choice of which solution type to focus on for the substantive interpretation will also be based upon the research aims of a given study, the complexity of the analysis, and the extent to which reasoning about logical remainders is possible in a given field of research. The next section examines the topic of logical remainders and counterfactual analysis, whereas Chapter 9 will look into further aspects of solution terms.

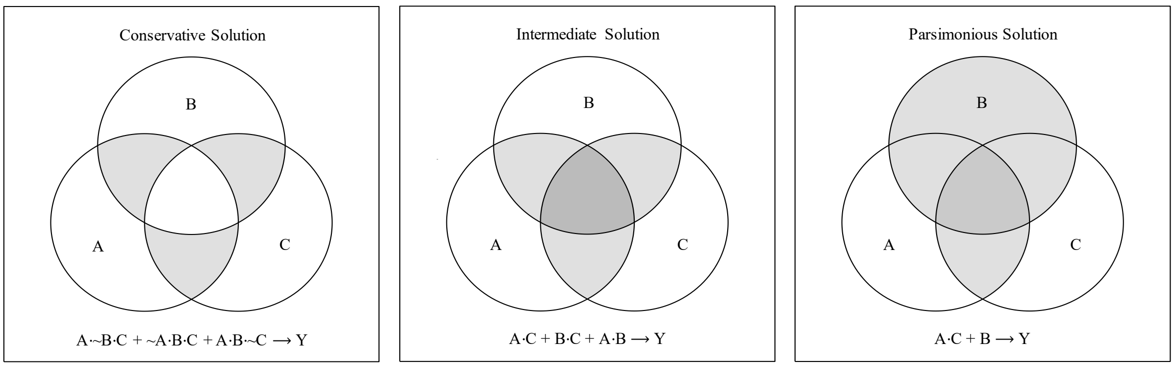
Table 7.7 Conservative, Intermediate, and Parsimonious Solution

Path		Relation	Consistency	PRI	Raw coverage	Unique coverage
Conservative solution			0.97	0.97	0.87	–
1	A·~B·C	+	0.95	0.95	0.37	0.29
2	~A·B·C	+	0.95	0.94	0.32	0.24
3	A·B·~C	⟶ Y	0.95	0.94	0.32	0.24
Intermediate solution			0.97	0.97	0.88	–
1	A·C	+	0.96	0.95	0.40	0.29
2	B·C	+	0.96	0.95	0.35	0.24
3	A·B	⟶ Y	0.95	0.95	0.35	0.24
Parsimonious solution			0.96	0.96	0.90	–
1	A·C	+	0.96	0.95	0.40	0.29
2	B	⟶ Y	0.96	0.95	0.61	0.51

Figure 7.3 The Complexity-Parsimony Dimension

A·~B·C +	A·C +	
~A·B·C +	B·C +	A·C +
A·B·~C	A·B	B
Complexity		*Parsimony*

Figure 7.4 Conservative, Intermediate, and Parsimonious Solution



While the layout of Table 7.7 is suitable for a comparison on the different solution terms, the Boolean notation is fairly abstract and may pose a challenge for readers who are not used to set-theoretic methods. Hence, for publications it is often a good idea to use a more reader-friendly layout where conditions are spelled out and case membership is listed, both of which enhances the communication of the analytical results (a topic we will return to in Chapter 10).

Table 7.8 shows an abbreviated example from a recent study by Tobias Ide, Miguel Rodriguez Lopez, Christiane Fröhlich, and Jürgen Scheffran (Ide et al. 2020). Their study on the conditions leading to water-related conflict in the Middle East and North Africa resulted in two pathways. Following the notation introduced by Charles Ragin and Peer Fiss (Fiss 2011; Ragin and Fiss 2008), black circles ('●') indicate the presence of a condition, whereas crossed-out circles ('⊗') refer to the absence of a condition. We can see of the four conditions included in the study, listed at the top left side of the table, the first path combined *cleavages* between social groups and an *autocratic regime*, whereas the second path contained *cleavages* in combination with *water cuts*. The condition *nightlight emissions* did not appear in any of the two solution paths. The center of the table entails information on measures of fit, including consistency, the raw coverage of each of the two paths (irrespective of empirical overlap), and their unique coverage. The bottom of the table lists cases by path membership. We can see that six cases hold membership in both solution paths (the non-underlined cases), while seven cases are underlined to indicate that they are uniquely covered by one of the two paths. The overall solution consistency and coverage are listed at the low end of the table.

This layout can be easily customized to include further information, for instance on *unexplained cases* not covered by the solution or on *deviant cases* that hold membership in the solution but not in the outcome (see Chapter 10). Presenting QCA results in this layout has the advantage that all of the condition names can be spelled out in full, while adding a visual element that makes it easier to grasp the combinations of conditions. Moreover, unless there is a very large number of cases, all of the covered cases can be listed by path membership and uniquely covered cases can be highlighted. Both of this enhances the connection between the abstract analysis and the case-oriented nature of QCA.

Table 7.8 Example: Solution Term and Solution Paths

	Paths	
	1	2
Cleavages	●	●
Autocratic regime	●	
Nightlight emissions		
Water cuts		●
Consistency	1.00	0.91
Raw coverage	0.53	0.59
Unique coverage	0.18	0.24
Covered cases /	AinBerda	Aidah
Uniquely covered cases	Annaba	AinBerda
(Underlined)	Batna	Annaba
	Damascus	Aramta
	Damru	Batna
	El Burullus	Damru
	ElChatt	Diyarbakir
	Guercif	El Burullus
	Ouargla City	El Chatt
		Nablus
Solution Consistency	0.93	
Solution Coverage	0.77	

Data source: Ide et al. (2020: 8).

Counterfactual Analysis

One of the strengths of QCA is that it uncovers limited diversity among social phenomena, which creates an opportunity for *counterfactual analysis* (Ragin 2008, 151). The truth table is the tool to identify configurations filled with empirical cases, rows that lead consistently towards the outcome, the ones that are inconsistent, and configurations for which there are no empirical cases – the logical remainder rows. Logical remainders are *potential* counterfactual cases. They are combinations of conditions without empirical instances for whom the plausibility of the outcome can be evaluated by the researcher.

As noted in the discussion of theories of causation (Chapter 4), counterfactual analysis has an established pedigree in the social sciences, where it infused both qualitative (e.g., Fearon 1991; Harvey 2012; Levy 2008; Tetlock and Belkin 1996) and quantitative approaches (e.g., Emmenegger 2011; King and Zeng 2007; Morgan and Winship 2007). In QCA, logical remainders are included in the parsimonious solution, but for this solution type the algorithm is in the driver's seat, because it selects solely on the basis of whether the remainders enable further minimization, aiming for the least complex solution. Hence, there is no counterfactual *reasoning* involved in the parsimonious solution.

The intermediate solution remedies this shortcoming because it allows researchers to engage in counterfactual reasoning for the incorporation of logical remainders. How is this to be done? The first criterion is the distinction between easy and difficult counterfactuals (Ragin 2008; Ragin and Sonnett 2005). *Easy counterfactuals* are simplifying logical remainders for which the presence of the outcome can be expected, based on theoretical and substantive knowledge. *Difficult counterfactuals* are simplifying assumptions that conflict with such knowledge. The conceptual distinction between easy and difficult counterfactuals thus emphasizes the importance of theoretical expectations and familiarity with the cases as preconditions for counterfactual reasoning (see also Mendel and Ragin 2011).

Yet, an applied researcher might ask *how* the distinction between easy and difficult counterfactuals ought to be made. On which grounds can we justify the use or non-use of logical remainders? Charles Ragin described the core idea behind simplifying assumptions concisely in *Fuzzy-Set Social Science* (2000), where he also stressed the need for transparent documentation:

> Ultimately, the selection of simplifying assumptions for incorporation into a causal generalization must be grounded as much as possible in theoretical and substantive knowledge. The purpose of each evaluation – that is, thought experiment – is to *assess the plausibility of a nonexistent combination of conditions as a simplifying assumption*. The researcher asks: If this combination of conditions existed empirically, would it generate the outcome? If the answer is "yes," then the simplifying assumption is permitted. Of course, if any simplifying assumptions are incorporated into a causal generalization, they should be carefully documented and thus made available for evaluation by the audience for the research (Ragin 2000, 305, emphasis added).

It is against this backdrop that subsequent studies elaborated the work on counterfactual reasoning and simplifying assumptions (Rihoux and De Meur 2009; Schneider and Wagemann 2012; 2013; Yamasaki and Rihoux 2009). Clearly, these and other works have advanced the discussion of logical remainders and enriched the analytical toolkit of QCA. Yet, this has also led to a virtual "logical remainder soup" with a plethora of related terms that can easily confuse new users, as well as those who are acquainted with set-theoretic methods. With that in mind, my aim in this section is to cut through the thicket and to provide a *concise* account of counterfactuals, their analytical potential, and the hazards of unwarranted assumptions.[14]

To begin with, there are three types of logical remainders that elude the distinction between easy and difficult counterfactuals. The first of these are *contradictory counterfactuals*, which are defined as logical remainders that are treated as if they were sufficient for both the outcome and the non-outcome (Rihoux and De Meur 2009, 64; Yamasaki and Rihoux 2009, 136).[15] Because the analysis of the outcome and the non-outcome are conducted separately, it may happen unwittingly that the same logical remainder is used to derive solutions for both of these. This would be a *logical contradiction* that should be solved, either by limiting the use of the logical remainder to the analysis where it is plausible (either for the outcome *or* the non-outcome), by excluding the logical remainder altogether, or by addressing the underlying issue through changes in the research design. The first two approaches would be technical solutions introduced during the analysis, whereas the third approach takes a step back to find ways how to improve the research design to prevent such contradictions (see Chapter 2).

The second type of logical remainders to avoid are *impossible counterfactuals*, which are remainders that are simply *not possible* by what we know about the physical or social world (Schneider and Wagemann 2012, 206). This can mean that the respective configuration is either generally impossible or that it simply has never happened in history (irrespective of whether it might happen in the future). A crude but frequently used example is the combination of the sets *pregnant people* and *non-women* – "pregnant men" – which will be empty due to its impossibility. As an example for the second type, imagine a study that seeks to explore the conflict involvement of UN member states. The conditions include *permanent membership* in the UN Security Council and *possessing nuclear weapons.* Though both of these can be considered relevant conditions, this design would create logical remainders that combine permanent UNSC membership with the *absence* of nuclear weapons – an impossibility at the time of writing, though it is conceivable that at some point in the future the UN Security Council might comprise permanent member states that do not possess nuclear weapons.

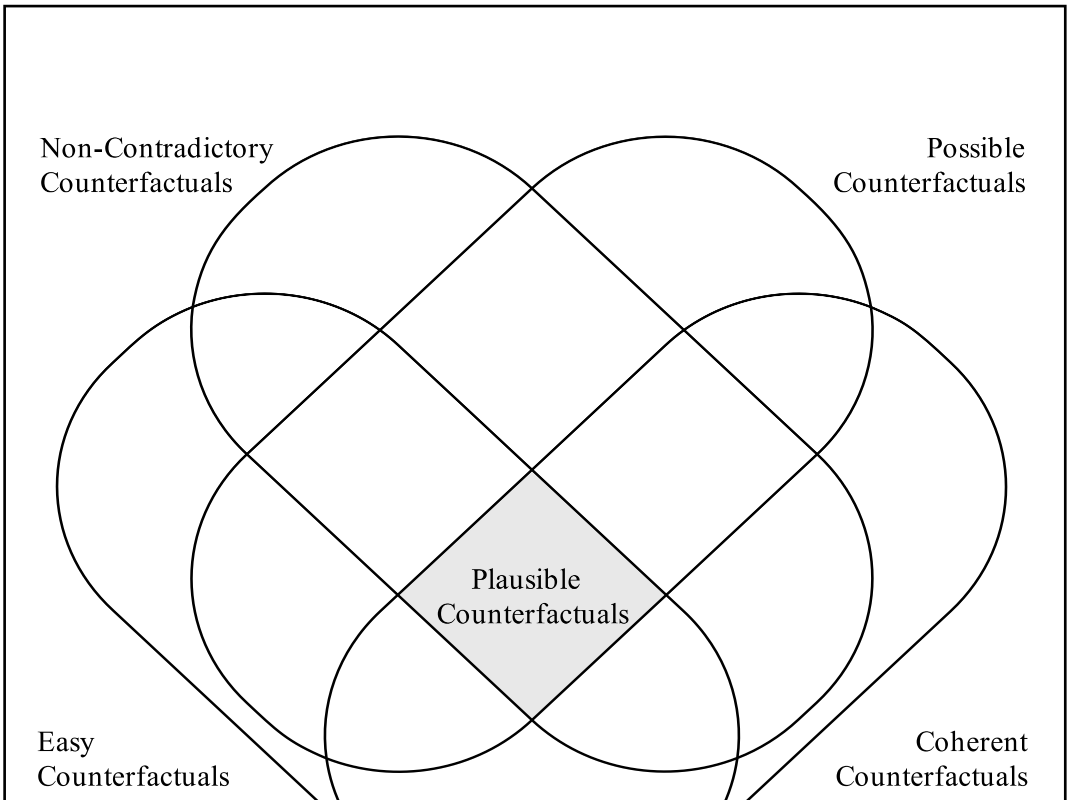
Impossible counterfactuals can occur whenever the selection of conditions allows for combinations that are not possible, but which would still be created as logical remainder rows in the truth table. Hence, such impossible counterfactuals may inform the parsimonious solution. Clearly, the best remedy to this situation is an improved research design. With QCA, only a limited number of conditions can be included and each of these should, in principle, be conceived in a way that allows for combinations with all other conditions. This mirrors the guidelines for case selection, where the *possibility* of a case showing the outcome is a crucial criterion (Mahoney and Goertz 2004). Wherever an adapted research design is not feasible, then the impossible counterfactuals should be excluded from the parsimonious and intermediate solutions.

Finally, the third type of logical remainders that should be avoided are *incoherent counterfactuals*, which are assumptions that conflict with a statement of necessity made earlier in the analysis (Schneider and Wagemann 2012, 201).[16] For example, we may have identified a relevant (non-trivial) necessary condition during the first step of the set-theoretic analysis (see Chapter 6). Now it could be seen as a logical contradiction if we treated remainders that do *not* include the necessary condition as sufficient for the outcome. Hence, any necessary condition that was identified during the first step of the analysis should constrain the set of logical remainders that are available for the parsimonious and intermediate solution. In principle, only those remainders that contain the presence of the necessary condition should be allowed.

However, this principle should not be followed blindly. In a given empirical setting, there may be logical remainders for which it can reasonably be argued on the grounds of theoretical and substantive knowledge that the outcome can be expected although these might not entail the necessary condition. At the least, this possibility should *not be ruled out* per se, without thoroughly investigating the remainders and assessing their plausibility.

What does this mean for the analysis? The essential point for the truth table analysis is that counterfactuals should be located in the shaded area at the center of Figure 7.5. I term these *plausible counterfactuals* which are those logical remainders that meet all of the following criteria: (1) theoretical and substantive knowledge suggests the presence of the outcome, (2) the respective configuration is not contradictory, (3) empirically possible, and (4) does not conflict with a statement of necessity.[17] Whenever researchers engage in counterfactual reasoning with QCA, they should ensure to only include plausible counterfactuals in the analysis. A discussion of how this is done with the software is provided in the *online material* for this book (see Appendix for instructions).

Figure 7.5 Types of Counterfactuals



Apart from the guidelines discussed above, two specific approaches have been developed to refine the treatment of logical remainders. The *enhanced standard analysis* (ESA) by Carsten Schneider and Claudius Wagemann (2012) is a procedure designed to avoid "untenable assumptions" based on contradiction, incoherence, or impossibility during the minimization of the truth table. This yields so-called *enhanced* parsimonious and intermediate solutions which tend to be more complex than those derived with the standard routine. The ESA procedure is detailed in Schneider and Wagemann (2012, 198-211) and Duşa (2019, 183-96). This approach is complemented by a second approach, the *theory-guided enhanced standard analysis* (TESA, Schneider and Wagemann 2012, 211), which specifically allows for the inclusion of non-simplifying remainders. The key point with TESA is that it emphasizes *theoretical resonance* over parsimony, and hence the produced solution terms will tend to be more complex than with the other approaches.

With their emphasis on tenability and analytical aims beyond parsimony, Schneider and Wagemann (2012; 2013) have advanced the methodological debate on the treatment of logical remainders in QCA. Notably, the "SetMethods" package for R has introduced a variety of advanced functions for set-theoretic analysis (Oana and Schneider 2018; Oana et al. 2021). Yet, the suggested analytical procedures for the treatment of logical remainders in ESA have also drawn critiques. Brian Cooper and Judith Glaesser (2016, 311) contend that ESA "goes beyond the available evidence" in ruling out logical remainder rows that do not show a previously identified necessary condition. Alrik Thiem (2016) holds that ESA effectively yields conservative solutions, due to the constraints imposed by previously identified necessary conditions. In their replies, Schneider and Wagemann emphasize the importance of meaningful criteria for the relevance of necessary conditions, underlining that ESA does *not* force researchers "to declare as necessary any condition that is a superset of the outcome" (Schneider 2018; Schneider and Wagemann 2016, 317). This underscores a point made throughout this book, namely that it is upon researchers to justify their analytical choices in the light of their case knowledge and theoretical expectations, which should not be replaced by mechanical routines.

Broadly conceived, counterfactual analysis is currently one of the most dynamic areas of QCA research – as it touches upon the linkage between theory and empirical analysis, the assumptions that go into solution terms, and the technical implementation of these procedures. Doubtless, this area will see further development as new procedures will be suggested for the treatment of logical remainders. We will revisit this topic in Chapter 9, when examining the current debate surrounding the validity of different solution terms in QCA.

Notes

[1] Goertz and Starr (2003, 3).

[2] Ragin (2014, xxvii).

[3] At some stage it has been recommended to *exclude* conditions that are found to be necessary from the following truth table procedure, while retaining them for the substantive interpretation (Ragin 2009, 110). However, removing necessary conditions from the analysis may decrease inferential leverage (Mello 2013, 13). Notably, Ragin shifted his position on this issue (Mendel and Ragin 2011, 24).

[4] The 0.9 consistency benchmark for necessary conditions is now widely accepted. Yet, it should be pointed out that in some of his earlier work, Ragin (2000, 229; 2003a, 194) applied the linguistic qualifiers known from fuzzy sets also to the analysis of necessary conditions (as in "more in than out" and other verbal qualifiers to characterize partial set membership, see Chapter 5). Hence, conditions with a consistency of at least 0.8 were considered "almost always" necessary for the outcome (Ragin 2003a, 194). For relevance of necessity, Schneider and Wagemann point out that "[l]ow values indicate trivialness and high values relevance" (2012, 237), and Oana et al. (2021, 205) suggest a threshold of 0.6, above which conditions may be deemed relevant.

[5] On omitted variables in QCA, see Radaelli and Wagemann (2019).

[6] The example has been simplified for the purposes of illustration. Among other differences, Mantilla (2012) uses fuzzy-set QCA and includes two observations for Peru (based on different time periods).

[7] Schneider and Wagemann (2012, 105) rightly underline that the removed condition should be considered "irrelevant" only under the presence of the remaining conditions.

[8] For a detailed technical discussion of the algorithmic differences, see Duşa (2019, 197-208).

[9] Exceptions may be permissible when consistency levels are generally low for a given truth table or when there is a respective row that carries a lot of empirical weight, and which comes near the 0.75 consistency threshold (for example, a row with 0.74 consistency with an important case that should be retained for the truth table analysis). However, rather than using a low threshold, I recommend revisiting the research design (Chapter 2) and the calibration of conditions (Chapter 5) in order to construct a more consistent truth table.

[10] Duşa's (2019, 177) argument departs from Schneider and Wagemann (2012, 166), who suggest that "the formula without any assumptions is not necessarily the most complex one".

Irrespective of these differences, it is apt to use the term *conservative* solution, as Schneider and Wagemann (2012, 162) propose, because this solution makes no counterfactual assumptions.

[11] The intermediate solution is at times reduced solely to these "directional expectations". However, Ragin apparently used the term only in a working paper (Ragin 2003b), whereas his many discussions of simplifying assumptions make no reference to it but convey a broader understanding that emphasizes *plausibility* as a key criterion (Ragin 2000; 2008; 2009). As Ragin notes (2008, 162), the distinction between easy and difficult counterfactuals "is not a rigid dichotomy, but rather a continuum of plausibility" that rests "primarily on the state of existing theoretical and substantive knowledge in the social scientific community at large". For broader conceptions of the intermediate solution in empirical applications, see, among many others, Maggetti and Levi-Faur (2013), Mello (2020), and Vis et al. (2013).

[12] The solution terms can occasionally be *identical*. Apart from the situation of a complete truth table without logical remainder rows (which happens rarely), this can occur when the logical remainders do not allow for further minimization or when the assumptions made for the intermediate solution effectively match those for the other solutions.

[13] The Venn diagrams serve to illustrate the relationship between the solution types, without taking into account degrees of fuzzy-set membership in the underlying data.

[14] To foster this aim, I have streamlined the terminology and reduced it to what I regard as essential differences between the concepts used by various authors. Readers are encouraged to consult the referenced sources for complementary accounts.

[15] These are also referred to as "contradictory simplifying assumptions" (Yamasaki and Rihoux 2009, 136) and subsumed under the category of "incoherent counterfactuals" (Schneider and Wagemann 2012, 198). To avoid ambiguity, I use the latter term only for remainders that conflict with necessary conditions.

[16] Ragin notes this problem in his conversation with Jerry Mendel, recommending the exclusion of logical remainders that do not show an identified necessary condition (Mendel and Ragin 2011, 25).

[17] I use the term *plausible counterfactuals* because it resonates closely with the core idea of "plausibility" in counterfactual reasoning (Ragin 2000, 300-08; 2008; 2009). A related but broader term is "good counterfactuals", which further entails non-simplifying logical remainders (Schneider and Wagemann 2012, 212).

References

Baumgartner, Michael, and Alrik Thiem. 2017. "Model Ambiguities in Configurational Comparative Research." *Sociological Methods & Research* 46 (4): 954-87.

Cooper, Barry, and Judith Glaesser. 2016. "Qualitative Comparative Analysis, Necessary Conditions, and Limited Diversity: Some Problematic Consequences of Schneider and Wagemann's Enhanced Standard Analysis." *Field Methods* 28 (3): 300-15.

Duşa, Adrian. 2018. "Consistency Cubes: A Fast, Efficient Method for Exact Boolean Minimization." *The R Journal* 10 (2): 357-70.

———. 2019. *QCA with R. A Comprehensive Resource*. Cham: Springer.

Emmenegger, Patrick. 2011. "How Good Are Your Counterfactuals? Assessing Quantitative Macro-Comparative Welfare State Research with Qualitative Criteria." *Journal of European Social Policy* 21 (4): 365-80.

Fearon, James D. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43 (2): 169–95.

Fiss, Peer C. 2011. "Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research." *Academy of Management Journal* 54 (2): 393-420.

Goertz, Gary, and Harvey Starr. 2003. "Introduction: Necessary Condition Logics, Research Design, and Theory." In *Necessary Conditions*, edited by Goertz, Gary and Harvey Starr. Lanham: Rowman & Littlefield, 1–23.

Harvey, Frank P. 2012. *Explaining the Iraq War: Counterfactual Theory, Logic, and Evidence*. New York: Cambridge University Press.

Ide, Tobias, Miguel Rodriguez Lopez, Christiane Fröhlich, and Jürgen Scheffran. 2020. "Pathways to Water Conflict During Drought in the MENA Region." *Journal of Peace Research* online first (doi: 10.1177/0022343320910777): 1-15.

King, Gary, and Langche Zeng. 2007. "When Can History Be Our Guide? The Pitfalls of Counterfactual Inference." *International Studies Quarterly* 51 (1): 183-210.

Levy, Jack S. 2008. "Counterfactuals and Case Studies." In *The Oxford Handbook of Political Methodology*, edited by Box-Steffensmeier, Janet M., Henry E. Brady and David Collier. Oxford: Oxford University Press, 627-44.

Maggetti, Martino, and David Levi-Faur. 2013. "Dealing with Errors in QCA." *Political Research Quarterly* 66 (1): 198-204.

Mahoney, James, and Gary Goertz. 2004. "The Possibility Principle: Choosing Negative Cases in Comparative Research." *The American Political Science Review* 98 (4): 653–69.

Mantilla, Luis Felipe. 2012. "Mobilizing Religion for Democracy: Explaining Catholic Church Support for Democratization in South America." *Politics and Religion* 3 (3): 553–79.

Mello, Patrick A. 2013. "From Prospect to Practice: A Critical Review of Applications in Fuzzy-Set Qualitative Comparative Analysis." 8th Pan-European Conference on International Relations, Warsaw, 18-21 September.

———. 2020. "Paths towards Coalition Defection: Democracies and Withdrawal from the Iraq War." *European Journal of International Security* 5 (1): 45-76.

Mendel, Jerry M., and Charles C. Ragin. 2011. "fsQCA: Dialog Between Jerry M. Mendel and Charles C. Ragin." *USC-SIPI Report* (411): 1-43.

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

Oana, Ioana-Elena, and Carsten Q. Schneider. 2018. "SetMethods: An Add-on R Package for Advanced QCA." *The R Journal* 10 (1): 507-33.

Oana, Ioana-Elena, Carsten Q. Schneider, and Eva Thomann. 2021. *Qualitative Comparative Analysis Using R: A Beginner's Guide*. New York: Cambridge University Press.

Radaelli, Claudio M., and Claudius Wagemann. 2019. "What Did I Leave Out? Omitted Variables in Regression and Qualitative Comparative Analysis." *European Political Science* 18 (2): 275-90.

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

———. 2000. *Fuzzy-Set Social Science*. Chicago, IL: University of Chicago Press.

———. 2003a. "Fuzzy-Set Analysis of Necessary Conditions." In *Necessary Conditions*, edited by Goertz, Gary and Harvey Starr. Lanham: Rowman & Littlefield, 179–96.

———. 2003b. "Recent Advances in Fuzzy-Set Methods and their Application to Policy Questions." *COMPASSS Working Papers* 2 (9): 1-33.

———. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.

———. 2009. "Qualitative Comparative Analysis Using Fuzzy Sets (fsQCA)." In *Configurational Comparative Methods*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 87–121.

———. 2014. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. 2nd ed. Berkeley, CA: University of California Press.

Ragin, Charles C., and Peer C. Fiss. 2008. "Net Effects Versus Configurations: An Empirical Demonstration." In *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, edited by Ragin, Charles C. Chicago, IL: University of Chicago Press, 190-212.

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

Rihoux, Benoît, and Gisèle De Meur. 2009. "Crisp-Set Qualitative Comparative Analysis (csQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 33-68.

Rubinson, Claude. 2013. "Contradictions in fsQCA." *Quality & Quantity* 47: 2847-67.

Rubinson, Claude, Lasse Gerrits, Roel Rutten, and Thomas Greckhamer. 2019. "Avoiding Common Errors in QCA: A Short Guide for New Practitioners."

Schneider, Carsten Q. 2018. "Realists and Idealists in QCA." *Political Analysis*: 1-9.

Schneider, Carsten Q., and Seraphine F. Maerz. 2017. "Legitimation, Cooptation, and Repression and the Survival of Electoral Autocracies." *Zeitschrift für Vergleichende Politikwissenschaft* 11 (2): 213-35.

Schneider, Carsten Q., and Claudius Wagemann. 2010. "Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets." *Comparative Sociology* 9 (3): 397–418.

———. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. New York, NY: Cambridge University Press.

———. 2013. "Doing Justice to Logical Remainders in QCA: Moving Beyond the Standard Analysis." *Political Research Quarterly* 66 (1): 211-20.

———. 2016. "Assessing ESA on What It Is Designed for: A Reply to Cooper and Glaesser." *Field Methods* 28 (3): 316-21.

Tetlock, Philip E., and Aaron Belkin, ed. 1996. *Counterfactual Thought Experiments in World Politics: Logical, Methodological, and Psychological Perspectives.* Princeton, NJ: Princeton University Press.

Thiem, Alrik. 2016. "Standards of Good Practice and the Methodology of Necessary Conditions in Qualitative Comparative Analysis." *Political Analysis* 24 (4): 478-84.

Vis, Barbara, Jaap Woldendorp, and Hans Keman. 2013. “Examining Variation in Economic Performance Using Fuzzy-Sets.” *Quality & Quantity* 47 (4): 1971-89.

Yamasaki, Sakura, and Benoît Rihoux. 2009. “A Commented Review of Applications.” In *Configurational Comparative Methods*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 123–45.