Chapter 3
Set Theory

Mello, Patrick A. (2021) *Qualitative Comparative Analysis: An Introduction to Research Design and Application,* Washington, DC: Georgetown University Press, Chapter 3.

[Q]ualitative research is often based, explicitly or implicitly, on set theory and logic, and these mathematical tools must be comprehended in their own right […].

Gary Goertz and James Mahoney [1]

QCA is a set-theoretic method. But what are *sets* and how does a *set-theoretic* approach differ from other kinds of approaches? At the most basic level, a set can be defined as a group of elements that share certain characteristics. A set is a class of objects (Quine 1969). For George Lakoff and Rafael Núñez, one way to see sets is as a "containerlike entity" (Lakoff and Núñez 2000, 140) and James Mahoney describes sets as "boundaries that define zones of inclusion and exclusion" (Mahoney 2010, 7). Given these understandings, it is apparent that there is a close affinity between sets and social science concepts. In fact, most social science theories are grounded in set theory, either explicitly by using the language of necessary and sufficient conditions, or implicitly by invoking the set-theoretic logic without expressly referring to necessity and sufficiency (Goertz and Mahoney 2012; Ragin 1987; Schneider and Wagemann 2012).[2]

This chapter lays out the set-theoretic foundations of QCA, including key terms and concepts on which later parts of the book will build. Starting with the distinctive elements of a set-theoretic approach as opposed to a statistical approach, the chapter introduces crisp and fuzzy sets, set operations, formal notation, and the functions of truth tables. The chapter closes with a discussion of necessary and sufficient conditions and the assessment of set relations.

Sets exist all across the social world. Social kinds of sets are concepts that we use to describe groups of objects with specific characteristics. For instance, we may seek to study democratic states or developed countries. Yet, the referent concepts of "democracy" and "development" are not objectively given but created in the human mind. As social science concepts, they may be "essentially contested" (Gallie 1956), and their meaning may have changed over time. We know that "democracy" means something else today than 100 years ago, when it was not seen as an essential property of democracy to have universal voting rights. Likewise, a country's "development" was long understood narrowly in terms of the size of its economy, whereas measures like the UN's Human Development Index rest on a broader conception that includes education, standard of living, and life expectancy (UN 2018).

What distinguishes a set-theoretic approach from other approaches? To illustrate this, we can juxtapose the set-theoretic approach with a statistical approach to derive six key differences (Goertz and Mahoney 2012; Ragin 2014, xxiii).[3] To begin with, the statistical approach seeks to explain the occurrence of a dependent variable (*explanandum*) with one or more independent variables (*explanans*). These variables are labelled using nouns, as in *height* or *partisanship*. For each of the included variables the selected cases are then scored based on measurement through indicators. Studies following this approach typically seek to identify linear relationships as in correlations between the independent variables and the dependent variable, summarized in a correlation matrix, with the aim of identifying the net effect of each independent variable on the dependent variable.

By contrast, the set-theoretic approach uses the terms *conditions* and *outcome* rather than independent variables and dependent variable. The different terminology serves to emphasize the *qualitative* and *case-oriented* nature of set-theoretic analysis, which requires careful concept formation before the actual set-theoretic analysis (Ragin 2014, xxv).[4] For each of these, nouns and adjectives are used. Hence the set-theoretic approach would refer to *tall* people or *left* partisanship. Other differences concern the way data is generated and analyzed. The set-theoretic approach refers to set membership scores which are bounded between 0 and 1, as the minimum and maximum values that can be assigned. By contrast, variable-oriented approaches typically work with unbounded numbers and various kinds of numerical scales can be used. Whereas operations on sets follow the rules of Boolean algebra and set theory – which will be introduced in this chapter – variables can be manipulated by applying the rules of linear algebra.[5]

Moreover, sets are *calibrated* towards a predefined point of reference – we could say that they are directed towards a qualitative state – and hence there is more information entailed in calibrated than in uncalibrated measures (Goertz 2020; Ragin 2008a; 2008b). This means that once we know which concept a set refers to, then we also know how to interpret a case's membership score in that set. For example, a person with a score of 1 in the set of *tall people* would be a tall person. Now consider a person with a *height* of 5 feet 10 inches. Whether that person should be considered tall or not will depend on contextual information like the person's gender, country of origin, and whether the data is based on a historical or contemporary record.

Finally, the set-theoretic approach aims to identify causally complex set relations involving necessary and sufficient conditions. The data is summarized in the *truth table*, from which solution terms or *causal recipes* are derived (Ragin 2014, xxvii). Table 3.1 juxtaposes these differences between the set-theoretic approach and the statistical approach. The bottom line is that it is vital to use the *correct terminology* when referring to each, because the terms are *not* synonyms but rather reflect distinct understandings of empirical research.

Table 3.1 Concepts in Statistical and Set-Theoretic Approaches

Differences by approaches	Set-theoretic approach	Statistical approach
Phenomenon to be explained (*explanandum*)	Outcome	Dependent variable
Phenomena to explain (*explanans*)	Conditions	Independent variables
Numerical conversion of concepts/raw data	Calibration	Measurement
Relationships to be explored	Causal complexity	Linear relationships
Analytical device	Truth table	Correlation matrix
Results	Necessary and sufficient conditions	Net effects of individual variables

Crisp and Fuzzy Sets

QCA is grounded in the works of George Boole, a British 19th century mathematician and logician whose *Laws of Thought* (Boole 1854) established the foundation for what was later termed Boolean algebra. Boole's approach uses variables that occur in only two states: *true* (present) or *false* (absent). This conception proved central to the development of electronic switching circuits and Boolean algebra was soon widely applied across the natural sciences (Whitesitt 2010), and also made inroads into the social sciences. Charles Ragin adopted these principles to develop a "Boolean approach" for comparative studies (Ragin et al. 1984), which later was labelled "Qualitative Comparative Analysis" (Ragin 1987). What is important for the aims of qualitative comparison is that Boolean algebra allowed for set-theoretic operations and the construction of truth tables – introduced in this chapter – as well as the minimization of truth tables to derive solution terms (discussed in Chapter 7).

The Boolean use of binary categories meant that QCA was, in its original form, limited to working with *crisp sets*, where 1 indicated the presence of a condition and 0 indicated its absence. Such a distinction emphasizes *qualitative* differences. Either a case belongs to a given set, or it does not. Yet, this procedure entails a loss of nuance, as regardless of how clear-cut our information on a case is, we must code it as either 1 or 0, and no further distinctions can be made (Rihoux and De Meur 2009).

The drawback of crisp-set QCA was overcome with the introduction of *fuzzy sets* (Ragin 2000), which allowed for graded set membership, as any scores from 0 to 1 became possible. Fuzzy logic was developed by Lotfi Zadeh (1965) as an extension of traditional set theory to tackle the problem of complex and imprecise concepts. Zadeh's work sparked a revolution in computer technology (McNeill and Freiberger 1993) and fuzzy sets have also made their way into the social sciences (Smithson and Verkuilen 2006), including linguistics (Lakoff 1973), and many other areas of application.

Fuzzy-set QCA combines qualitative and quantitative dimensions. Based on substantive and theoretical knowledge of their topic, researchers establish three *empirical anchors* that determine whether a case is considered to be "fully in" a given set (reflected in a fuzzy score of 1), whether it is "fully out" a given set (fuzzy score of 0), or whether it is "neither in nor out" a given set (fuzzy score of 0.5). The latter is the cross-over, or "point of maximum ambiguity", which means that based on the available evidence it is not possible to say whether the case is rather inside or outside a respective set (Ragin 2000; 2008b; Ragin and Fiss 2017). The empirical anchors form the basis for the calibration procedure, covered in Chapter 5.

Fuzzy sets clearly were a major step forward in the evolution of QCA. They allowed for fine-grained differentiation and – addressing concerns of some of the method's critics – showed that QCA did not have to be rooted in a deterministic understanding of causality. They also paved the way for *measures of fit* that enabled researchers to assess the quality and robustness of their analytical results (Chapter 6). Measures of fit established benchmarks to distinguish "perfect" necessary and/or sufficient conditions from "almost" necessary and/or sufficient conditions (those that do not meet a deterministic criterion of necessity or sufficiency), and to identify situations where there is no set-theoretic relationship. Finally, fuzzy sets eased the transformation of quantitative raw data into set membership values, through software-based procedures of calibration.

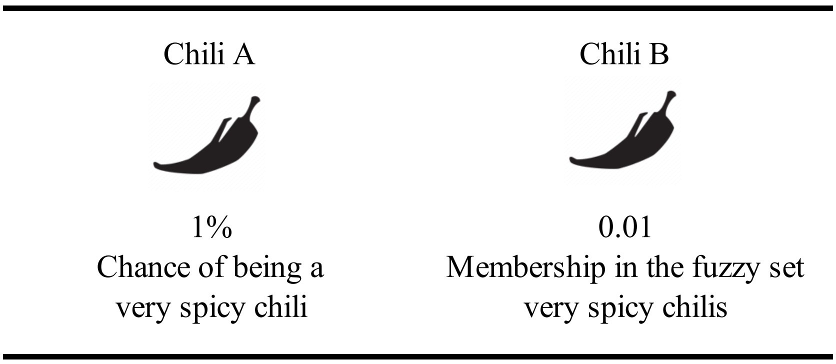
Irrespective of these advantages of fuzzy sets, it is important to note that the Boolean logic still applies with fuzzy sets – cases are treated as either inside or outside of a given set and the differences in degree only indicate that they are *more or less* in or out of those sets. The researcher still has to determine a criterion on how to distinguish cases that are rather inside a set from those that are rather outside.

This challenge has been described as Sorites Paradox or the paradox of the heap (derived from the Greek word *sorós*). Imagine a heap of sand. Now, we may assume that the heap comprises 1,000,000 grains of sand. Removing a single grain of sand will not change its character as a heap, since 999,999 grains of sand are still a heap of sand. And so it will be if another grain of sand is removed, and another, and so forth. Yet at some point we will be left with a few grains of sand that are clearly not a heap anymore. This prompts the question at which point the heap has turned into a *non-heap* – how many grains did we have to remove until the heap disappears?

To solve the paradox, we need to determine criteria to distinguish one state from the other. This is challenging for the heap of sand, because any firm numerical threshold will appear arbitrary. For social science concepts it can also be difficult, but it is a helpful practice that lets us think about *difference-makers* and the essential characteristics of our concepts.

A common misconception about fuzzy sets is that these somehow reflect probabilities. However, that is not the case. Let us take an example to illustrate the difference (Figure 3.1). Suppose you have two red chilis of identical appearance. You have no other information but that the first chili has a 1% chance of being a very spicy chili, whereas the second chili has a fuzzy score of 0.01 membership in the set *very spicy chilis*. You want to cook a non-spicy dish with mild chili as an ingredient. Which of these chilis would you choose? At first glance, it may appear that there is no difference between the two options. However, on closer look it should become clear that for Chili A the odds are 1 out of 100 that you will pick a very spicy chili. For Chili B the information given is not a probability – you *know for your sure* that Chili B is almost entirely outside the set of very spicy chilis – so this would be the better choice in the example. Fuzzy sets thus attribute a discrete score to a specific case and there is no uncertainty or probability involved.

Figure 3.1 Probabilities and Fuzzy Sets



Set Operations

Because QCA rests on Boolean algebra, we can use its rules for operations on sets to systematically assess the relationship between our outcome, conditions, and combinations of conditions. For our purposes, three operations from Boolean algebra are relevant.[6] The logical operator AND refers to the *intersection* or, to use the term from propositional logic, *conjunction*, between sets. The logical OR refers to the *union* or *disjunction* between sets, whereas the logical NOT describes the *negation* or *complement* of a set.

What can be confusing about QCA is that it draws on several different notational systems. Depending on a subfield's custom, it may be preferred to use symbols and terms from Boolean algebra, the logic of propositions, or formal set theory. Clearly, the existence of several different terms for the same set operations complicates things. In this book, an effort is made to keep formal notation to a minimum and to consistently use the same notation throughout. To enhance the translation of terms, Table 3.2 summarizes the operators and notational forms that are commonly used, depending on whether the reference point is Boolean algebra, set theory, or propositional logic.[7]

Table 3.2 Logical Operators and Notational Systems

Logical Operator	Boolean Algebra	Set Theory	Propositional Logic
AND	Multiplication $A \cdot B$	Intersection $A \cap B$	Conjunction $A \wedge B$
OR	Addition $A + B$	Union $A \cup B$	Disjunction $A \vee B$
NOT	Negation $1 - A$	Complement $\sim A$	Negation $\neg A$

How do set operations work in practice? The logical operators are best explained with examples. Table 3.3 gives hypothetical data for three cases, two crisp sets (labelled A and B), two fuzzy sets (C and D), and the results of calculations using the Boolean operators AND, OR, and NOT, in the six columns on the right-hand side of the table.

To begin, let us assume we are interested in cases that share two features. Only if both of these are present do we expect to see our outcome. In set-theoretic terms the *intersection* between two sets is expressed as $A \cdot B$, where $\cdot$ stands for AND (not to be confused with the mathematical multiplication sign). Verbally, the intersection reads as "A AND B". For convenience, we may also omit the operator and simply write the letters for the respective sets next to each other. To calculate a case's set membership in the intersection we take the *minimum* score across the sets, reflecting the weakest link or lowest common denominator between the respective sets. This means that Case 1 and Case 3 from Table 3.3 both receive a score of 0 in the expression $A \cdot B$.

For example, we may be interested in calculating two cases' memberships in the intersection between the sets *populist leader* (P) and *supportive public* (S). The first case has fuzzy scores of 0.9 (P) and 0.3 (S), whereas the second case has fuzzy scores of 0.6 (P) and 0.7 (S). Based on the calculation rule, the cases thus hold respective fuzzy scores of 0.3 and 0.6 in the intersection $P \cdot S$. The set-theoretic intersection resonates with settings were several features all have to be present and a high score in one condition cannot outweigh a low score in another.

We may also be interested in a situation where cases show at least one of two features, either of which we expect to lead towards our outcome, but the cases may also show both features. The *union* between two sets is expressed as A + B, where + stands for OR. Verbally, the union thus reads as "A OR B". This operator is based on an *inclusive* conception of OR, as in "one or the other, or both". Again, this should not be confused with the mathematical operator for addition. Accordingly, a case's set membership in the expression A + B is the *maximum* score across the two sets.

It follows that Case 1 and Case 2 from Table 3.3 get the same score for A + B, because each case has at least one set with a score of 1 and the maximum score determines the overall value. As an example, we may stipulate that governments that face either *constitutional restrictions* (C), *legislative veto players* (L), or both of these, will refrain from power abuses. The first case holds fuzzy scores of 0.7 (C) and 0.3 (L), whereas the second holds fuzzy scores of 0.2 (C) and 0.7 (V). Despite these differences, based on the calculation rule, both of these cases would receive fuzzy scores of 0.7 in the union C · L. The set-theoretic union reflects situations where there are multiple equivalent factors and either of them suffices for the outcome to occur.

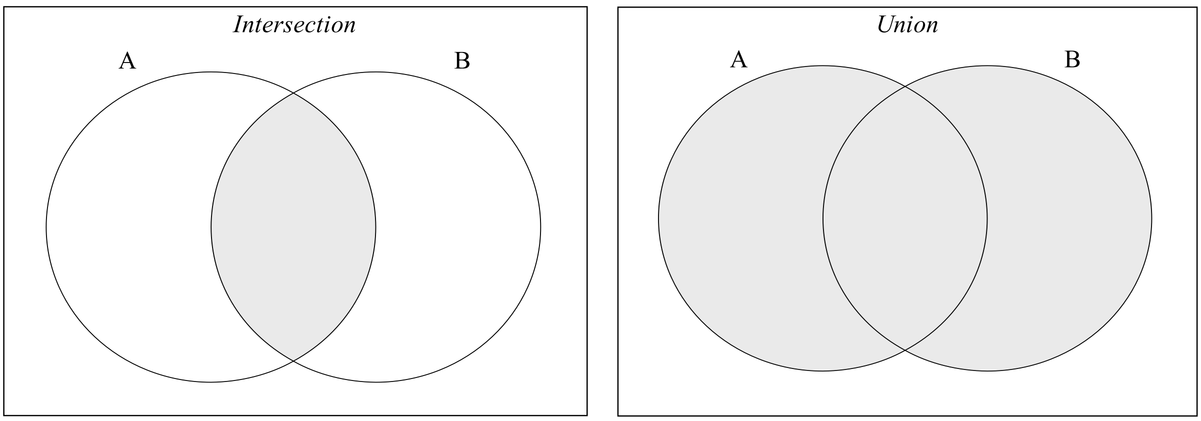
The final set-theoretic operator refers to the *negation* of a set. Once we know that a case holds membership in a set, we can also calculate its membership in the *non*-set, or the negation of the set. Formally, this is calculated as 1 – A. Here, the minus sign actually refers to the mathematical operator of subtraction. The last two columns of Table 3.3 show the values for non-A and non-C. Note that due to the bounded nature of set values, the scores for a condition and its negation will always add up to 1. For instance, Case 1 has a fuzzy-set membership of 0.9 in C and a fuzzy-set membership of 0.1 in ~C, which adds up to 1. For example, Germany may hold a fuzzy score of 0.2 in the set *high unemployment*, which means that it has a fuzzy score of 0.8 in the set *not*-high unemployment (1 – 0.2). When working with the negation of a set, it is vital to clarify the conceptual basis of the first set, before interpreting scores in the negation of that set. Some concepts may be symmetrical, but many are asymmetric in nature. We return to this aspect in Chapter 5.

Table 3.3 Boolean Operations with Crisp and Fuzzy Sets

	Crisp Sets		Fuzzy Sets		Boolean AND		Boolean OR		Boolean NOT	
	A	B	C	D	A · B	C · D	A + B	C + D	~ A	~ C
Case 1	1	0	0.9	0.3	0	0.3	1	0.9	0	0.1
Case 2	1	1	0.7	0.8	1	0.7	1	0.8	0	0.3
Case 3	0	0	0.2	0.4	0	0.2	0	0.4	1	0.8

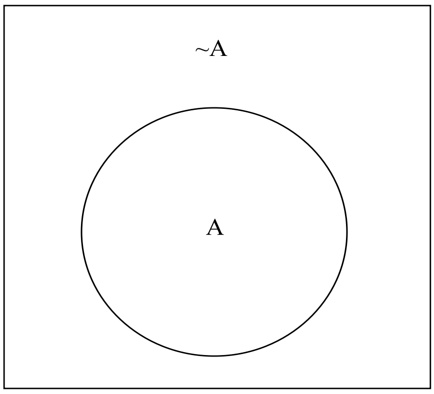
Set-theoretic operations and relationships between crisp sets can be effectively visualized with Venn and Euler diagrams (Goertz and Mahoney 2012; Mahoney and Sweet Vanderpoel 2015; Quine 1982; Rubinson 2019; Schneider and Wagemann 2012). In *Venn diagrams*, overlapping circles inside a rectangle depict all logically possible intersections between the respective sets. This distinguishes them from *Euler diagrams*, which show only the empirically existing intersections (Shin 1994). Both types of diagrams can be helpful to illustrate set relations. The following figures show the previously discussed Boolean operations in simple settings with just two sets. Clearly, one could add further circles for additional sets, which can be a useful heuristic device when thinking about the relationship among explanatory conditions in a given QCA model. Venn diagrams are also useful to visualize the configurations of truth table rows, discussed in the next section, because the number of areas inside the rectangle always matches the number of logically possible configurations in a given QCA model (Duşa 2019, 263). The left panel in Figure 3.2 shows the intersection of sets A and B (logical operator AND), whereas the right panel displays their union (logical operator OR).

Figure 3.2 Intersection and Union of Two Sets



Apart from the Boolean operators, there are two particular kinds of sets, which can become important in set-theoretic operations (Schneider and Wagemann 2012, 48-49). As its name implies, the *empty set* has no elements. It is designated by the symbol Ø and it will result from the intersection of a set with its complement (or negation). This is so because a case cannot at the same time be a member of a given set and its negation, as illustrated in Figure 4.4. Every case in the figure must be either a member of A or of ~A, as it cannot hold membership in both. Formally, this is expressed as A · ~A = Ø. This rule applies to crisp sets. With fuzzy sets, a case can have *partial* membership in a set and its negation, yet it can never have scores greater than 0.5 in both of these (Schneider and Wagemann 2012, 49). By contrast, the *universal set* is the set of all elements. As such, every other set is a subset of the universal set, designated by *U* (Smithson and Verkuilen 2006, 5). In Figure 3.3, the universal set describes everything that is inside the square box, which is the set A and its complement ~A. Formally, this is expressed as A + ~A = *U*.

Figure 3.3 Set A and Its Complement



Truth Tables

A central analytical tool for QCA is the *truth table*. The analytical procedure will be discussed in detail in Chapter 7, but for now it suffices to introduce the key features of truth tables and to describe how they differ from a common data spreadsheet.

The truth table shows the number of logically possible combinations of conditions included in a study. The term stems from formal logic where truth values for propositions are listed in table form (McCawley 1993; Tomassi 1999). To be sure, there is no sort of "higher truth" involved in truth tables! Each row in the truth table refers to a specific combination of conditions (a *configuration*). The number of rows equals the total number of possible configurations. The formula for this is 2^k, where k is the number of conditions included. This means that for two conditions there will be four rows, three conditions will result in 8 rows, four conditions will yield 16 rows, and so forth. Individual cases are assigned to the row where they hold the highest membership score (Ragin 1987, Ch. 6).

Table 3.4 shows a simple example with just two conditions, an outcome, and 10 hypothetical cases that distribute across the rows of the truth table. We can see from the table that three rows are populated with cases, whereas one row contains no cases. Hence, there is a question mark in the outcome column, indicating that there is no empirical foundation on which to decide whether the respective configuration should be associated with the presence or the absence of the outcome. With more conditions, truth tables become exponentially larger and the likelihood increases that cases will not fill all of the logically possible combinations. This is the issue of *limited diversity*, addressed in Chapter 2 in the context of the selection of cases and conditions, and to which we will return when discussing the treatment of logical remainders in Chapter 7.

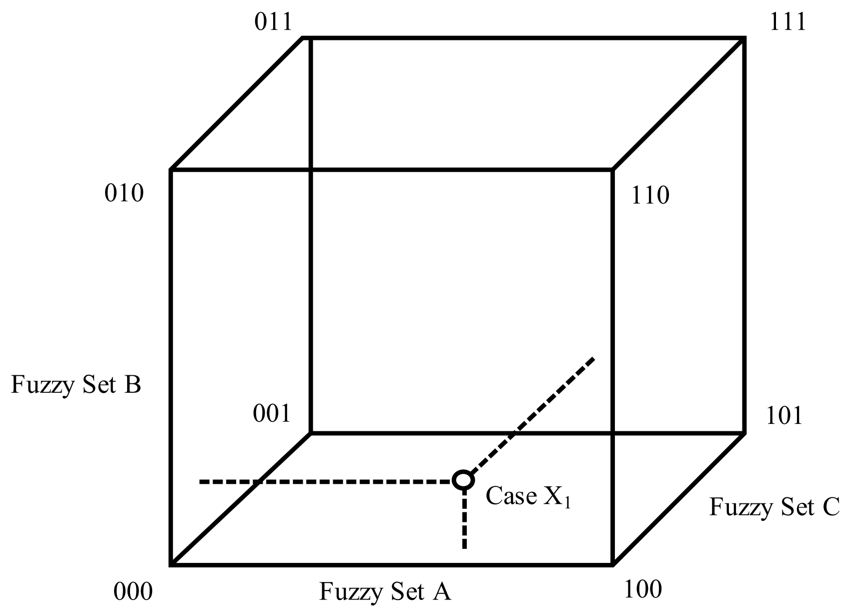
We can also see that the combinations represented by the top two rows are associated with a positive outcome, whereas row three is associated with a negative outcome. Based on the empirical cases, we can say that A · B (Row 1) and A · ~B (Row 2) lead to the outcome Y. This kind of information forms the basis for the *truth table analysis*, where a minimization algorithm will be applied (via software) to reduce the Boolean expressions and gain more parsimonious solution terms for the explanation of the outcome.

Table 3.4 A Simple Truth Table

Conditions		Outcome	Number of Cases
A	B	Y	
1	1	1	5
1	0	1	3
0	1	0	2
0	0	?	–

From a conceptual point of view, once fuzzy sets are introduced, we can conceive of the truth table as a multi-dimensional "attribute space" (Lazarsfeld 1937, 138), which circumscribes an area with as many dimensions as the number of fuzzy sets that are entailed in the analysis. Based on their fuzzy set membership scores, cases can be located at distinct points within the attribute space (Ragin 2000, Ch. 7). This is shown in Figure 3.4 for an analysis with three fuzzy sets. The key point is that cases can be anywhere inside the attribute space, whereas ideal types – with full membership or full non-membership in the respective sets – are located in the corners of the attribute space (from 000 in the lower left to 111 in the upper right corner). This feature inspired the QCA variant of fuzzy set ideal type analysis (Kvist 2007), where empirical cases are set in relation to pre-conceived ideal types (see Chapter 8).

Figure 3.4 Truth Table as a Multi-Dimensional Attribute Space



Necessary and Sufficient Conditions

As a set-theoretic method, the analytical procedure of QCA aims at the identification of necessary and sufficient conditions.[8] For Gary Goertz and James Mahoney (2012, : 12), the logic of necessity and sufficiency is what characterizes the heart of *qualitative* research.

A *necessary condition* means that the condition is always present when the outcome of interest occurs. Put differently, the outcome does not happen without the presence of the necessary condition. From a theoretical perspective, we can say that necessary conditions *explain failure* because they are a prerequisite for a phenomenon to occur.[9] For example, you may find that countries that experienced democratic breakdown always did so under conditions of economic crisis. This does not mean that an economic crisis causes a democratic breakdown in itself (as there may be other factors contributing to that outcome), but whenever there has been a breakdown of democracy it was preceded by an economic crisis.

A *sufficient condition* means that whenever the condition is present, the outcome occurs. For instance, a government change may be a sufficient condition for policy change. This does not mean that policy change can only occur after a government change (as there may be a variety of other factors that also lead to policy change), but for a certain population of cases it may be that whenever a new government has come to power, there has also been a subsequent policy change. Hence, we can say that sufficient conditions serve to *explain success*, as their presence always leads towards the outcome.

By conventional notation, a necessary condition is indicated by a left-pointing arrow, directed from the outcome towards the condition that is necessary: A ← Y, whereas a sufficient condition is indicated by a right-pointing arrow, directed from the condition that is sufficient towards the outcome: B → Y (Rihoux and Ragin 2009; Rubinson 2019; Schneider and Wagemann 2012). It is important to note that this notation differs from a common understanding of causal arrows, where the arrow runs from cause to effect. With a necessary condition, the left-pointing arrow rather indicates a dependency or implication, as in "Y implies A" or "no Y without A".

While the connection is often not made explicit, many social science theories are based on an understanding of necessary and sufficient causation (Goertz and Starr 2003; Ragin 1987). Here are some examples from studies in international relations, political science, and economics:

> If not a necessary condition, nuclear deterrence may be interpreted as a *sufficient condition* for peace (Gleditsch 1995, 543, original emphasis).

> The introduction of universal suffrage *led almost everywhere* (The United States excepted) to the development of Socialist parties (Duverger 1954, 66, emphasis added).

> One cannot have the productivity of an industrial society with political anarchy. But while such a state is *a necessary condition* for realizing the gains from trade, it obviously is *not sufficient*. (North 1984, 259, emphasis added).

> Thus, international pressure was a *necessary condition* for these policy shifts. On the other hand, *without domestic resonance*, international forces *would not have sufficed* to produce the accord, no matter how balanced and intellectually persuasive the overall package. (Putnam 1988, 430, emphasis added).

> In the rationalist perspective, however, a community of basic political values and norms is *at best a necessary condition* of [EU] enlargement [...] By contrast, in the sociological perspective, sharing a community of values and norms with outside states is *both necessary and sufficient* for their admission to the organization. (Schimmelfennig 2001, 61, emphasis added).

What these five quotes illustrate – and many more examples could be given – is that authors frequently base their arguments on a set-theoretic logic. The first quote from Nils Petter Gleditsch about nuclear deterrence is straightforward in its mention of a condition that is sufficient but not necessary. The second quote from Maurice Duverger about electoral rules differs from the first because it does not use the language of necessary and sufficient conditions.

However, its essence is set-theoretic: universal suffrage is seen as an (almost) sufficient condition for the development of Socialist parties. As such, Duverger's quote is a typical example of the implicit usage of set-theoretic language. The third quote from Douglass North turns the perspective by highlighting state institutions as a necessary condition for economic productivity. North also points out that while this condition is necessary, it is not sufficient to bring about the outcome. Similarly, Robert Putnam's quote focuses on international pressure as a necessary condition for policy shifts. Interestingly, Putnam also argues in combinatorial ways, emphasizing that it needed international pressure *and* domestic resonance, which together form a sufficient condition for policy shifts. Finally, the quote from Frank Schimmelfennig describes different expectations concerning necessity and sufficiency, based on competing theoretical perspectives. While rationalist theory expects a weak necessary condition, social constructivism expects a condition that is necessary and sufficient.

Assessing Set Relations

Now that we have defined necessary and sufficient conditions, what is missing is how we can identify these in empirical data. Chapter 6 will introduce several metrics to calculate set relations, leading up to the set-theoretic analysis that will be covered in Chapter 7. For now, we will examine three ways to *visualize* relations of necessity and sufficiency and to identify the respective data patterns. These include 2x2 tables (Figure 3.5), Euler diagrams (Figure 3.6), and XY plots or scatterplots (Figure 3.7). The figures are grouped together at the end of this section, to enhance their comparison.

An intuitive way to explore set-theoretic relationships for crisp-set data are 2x2 tables (Goertz and Mahoney 2012; Schneider and Wagemann 2012). Looking at the distribution of empirical cases across the four cells of a 2x2 table allows us to see whether there may be a set-theoretic relationship of necessity and/or sufficiency. Figure 3.5 shows separate 2x2 tables for relations of necessity and sufficiency. The left table identifies the cells that are relevant for a necessary condition. For necessary conditions, our focus rests on the presence of the outcome, indicated by the top row in the 2x2 table (the shaded area). If there are many cases that show the outcome and the condition (top right), and no cases that show the outcome but do not show the condition (top left), then this suggests that we have identified a necessary condition. Note that cases without the outcome are not important for the analysis of necessary conditions, though it is fine if there are some cases where the condition occurs without the outcome (bottom right). This merely indicates that the condition is not also sufficient for the outcome.

For sufficient conditions, the procedure is similar, but inverted. The right table in Figure 3.5 shows the cells that are relevant for a sufficient condition. Here, we direct our attention to the presence of the condition, indicated by the right-hand column (shaded area). When there are many cases that show the condition and the outcome, and no cases that show the condition but do not show the outcome, then this designates a sufficient condition. Cases without the condition are not important for the analysis of sufficient conditions, but it is unproblematic if there are some cases that show the outcome despite the absence of the condition, which simply means that the condition is not also necessary for the outcome.

These crisp-set relationships can also be visualized with a Euler diagram, as mentioned earlier in this chapter. Figure 3.6 displays perfect set relations of necessity (left panel) and sufficiency (right panel). In formal terms, a necessary condition is a *superset* of the outcome, which means that the circle for the condition X_1 fully encloses the outcome set Y. Hence, every case that holds membership in Y also holds membership in X_1. However, the reverse is not true, as can be seen from the dissimilar size of the circles, where many cases hold membership in the set X_1 but not in Y. There is also an area outside of both sets, representing cases that hold neither membership in the condition nor the outcome.

The Euler diagram on the right side of Figure 3.6 shows the relationship of sufficiency. Here the subset-superset relationship is inverted: a sufficient condition is a *subset* of the outcome, depicted in a circle for the condition X_2 that is fully enclosed by the outcome set Y. Every case that holds membership in X_2 also holds membership in Y. Yet, this does not account for all instances of Y, because there is a large area that is not covered by X_2 (but which may be accounted for by other conditions).

So far, we have only looked into examples with binary data, where cases are either inside or outside a set. What about fuzzy sets? For these, neither XY plots nor Venn or Euler diagrams will work because these visualizations cannot capture graded set membership. The best way to display fuzzy sets are XY plots, more generally referred to as *scatterplots*, which show the location of cases on a bivariate plot, with the condition on the x-axis and the outcome on the y-axis.

The left XY plot in Figure 3.7 shows a data pattern for a perfect necessary condition. We can see that all the empirical cases (shown in black dots) hold values for the outcome Y that are lower or equal to their respective values for the condition X_1. Hence, they reflect the expected set-theoretic relationship where the outcome Y is a *subset* of the condition X_1. This is highlighted by the diagonal line. The line separates cases with values that are equal to or higher for the outcome than for the condition ($Y \geq X_1$) from those that are equal to or lower for the outcome than for the condition ($Y \leq X_1$). Cases with equal values for the condition and the outcome are situated exactly on the diagonal line. Cases inside the lower gray triangle fulfill the criteria for perfect set relations of necessity.

For sufficient conditions, the situation is inverted. The right XY plot in Figure 3.7 displays the pattern for a perfect sufficient condition. All the empirical cases hold values for the outcome Y that are higher or equal to their respective values for the condition X_2, in line with the expected set-theoretic relationship where Y is a superset of the condition X_2. Hence, all of the cases are located in the upper gray triangle, above the diagonal line.

Figure 3.5 Necessary Condition and Sufficient Condition (2x2 Tables)

Necessary condition

Outcome	0	1
1	No cases	Many cases
0	Not important	Some cases

Sufficient condition

Outcome	0	1
1	Some cases	Many cases
0	Not important	No cases

Figure 3.6 Necessary Condition and Sufficient Condition (Euler Diagrams)

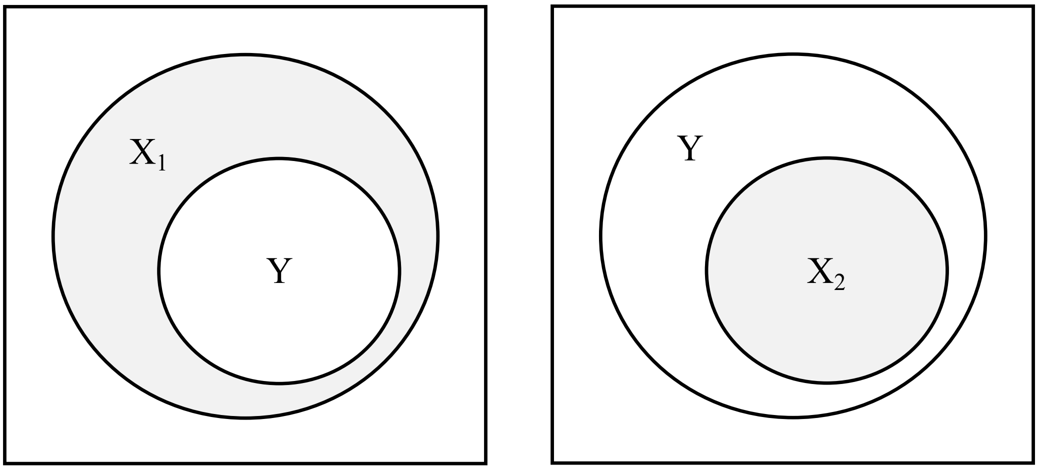
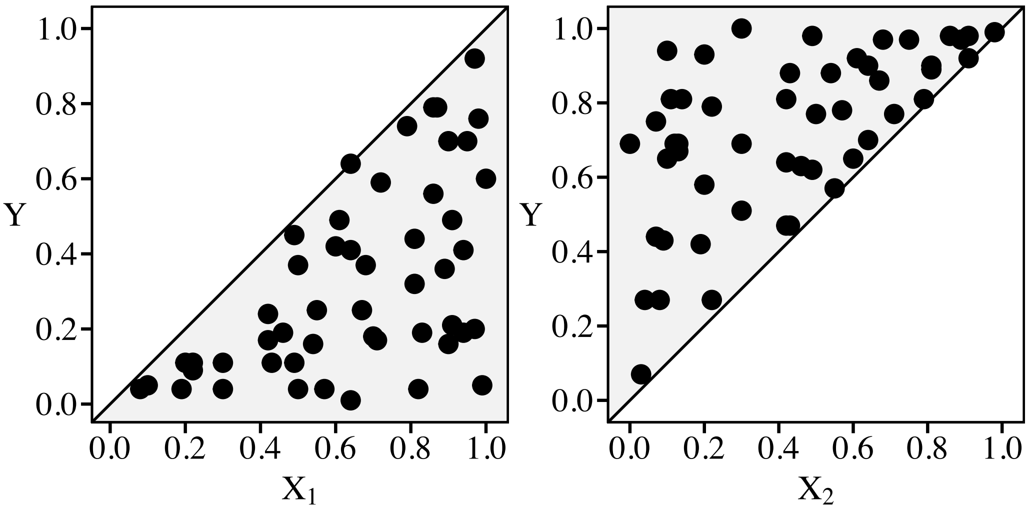


Figure 3.7 Necessary Condition and Sufficient Condition (XY Plots)



Notes

[1] Goertz and Mahoney (2012, 17)

[2] For an inventory of 150 necessary condition hypotheses, see Goertz (2003).

[3] This ideal-typical comparison should not imply that all statistical studies follow the described template.

[4] An empirical illustration of the conceptualization of the outcome and explanatory conditions in QCA is provided in Mello (2017, 130-33).

[5] Yet, set theory can be regarded as a "unifying theory for mathematics" (Cunningham 2016, ix).

[6] For comprehensive treatments of Boolean algebra and set theory, see Quine (1969; 1982), Potter (2004), Whitesitt (2010), and Cunningham (2016).

[7] Not listed here are two further operators that may be convenient depending on the theoretical context: The first is the *exclusive OR* (Hackett 2015; 2016), which describes membership in one or the other set, "but not both" as Hackett highlights (see also Quine 1982, 11-14). Another operator is the *set difference* between sets A and B, which is expressed as A \ B, or "A minus B" (Cunningham 2016, 4). For summaries of notational systems, see Quine (1982, Part I), Smithson and Verkuilen (2006, 6), Whitesitt (2010, 4), Schneider and Wagemann (2012, 54), Cunningham (2016, 6), and Rubinson (2019, 4).

[8] While QCA can be used for different purposes, as discussed in Chapter 2, many studies conduct some form of theory-guided analysis aimed at the identification of necessary and sufficient conditions (Mello 2013), which resonates with QCA's "orientation towards causal inference" (Wagemann 2017, 11). This topic is revisited in Chapter 4.

[9] Thanks to Gary Goertz for underlining this aspect.

Mello, Patrick A. (2021) *Qualitative Comparative Analysis: An Introduction to Research Design and Application,* Washington, DC: Georgetown University Press, Chapter 3.

References

Boole, George. 1854. *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*. Urbana, Illinois: Project Gutenberg.

Cunningham, Daniel W. 2016. *Set Theory: A First Course*. New York: Cambridge University Press.

Dușa, Adrian. 2019. *QCA with R. A Comprehensive Resource*. Cham: Springer.

Duverger, Maurice. 1954. *Political Parties: Their Organization and Activity in the Modern State*. London: Methuen.

Gallie, William B. 1956. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society* 56: 167–98.

Gleditsch, Nils Petter. 1995. "Democracy and the Future of European Peace." *European Journal of International Relations* 1 (4): 539-71.

Goertz, Gary. 2003. "The Substantive Importance of Necessary Condition Hypotheses." In *Necessary Conditions*, edited by Goertz, Gary and Harvey Starr. Lanham: Rowman & Littlefield, 65–94.

———. 2020. *Social Science Concepts and Measurement*. Princeton: Princeton University Press.

Goertz, Gary, and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Goertz, Gary, and Harvey Starr, ed. 2003. *Necessary Conditions: Theory, Methodology, and Applications.* Lanham: Rowman & Littlefield.

Hackett, Ursula. 2015. "But Not Both: The Exclusive Disjunction In Qualitative Comparative Analysis (QCA)." *Quality & Quantity* 49 (1): 75-92.

———. 2016. "The Goldilocks Principle: Applying the Exclusive Disjunction to Fuzzy Sets." *International Journal of Social Research Methodology* 19 (5): 551-74.

Kvist, Jon. 2007. "Fuzzy Set Ideal Type Analysis." *Journal of Business Research* 60 (5): 474–81.

Lakoff, George. 1973. "Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts." *Journal of Philosophical Logic* 2 (4): 458–508.

Lakoff, George, and Rafael E. Núñez. 2000. *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being*. New York: Basic Books.

Lazarsfeld, Paul. 1937. "Some Remarks on the Typological Procedures in Social Research." *Zeitschrift für Sozialforschung* 6 (1): 119-39.

Mahoney, James. 2010. "After KKV: The New Methodology of Qualitative Research." *World Politics* 62 (1): 120–47.

Mahoney, James, and Rachel Sweet Vanderpoel. 2015. "Set Diagrams and Qualitative Research." *Comparative Political Studies* 48 (1): 65-100.

McCawley, James D. 1993. *Everything that Linguists Have Always Wanted to Know about Logic But Were Ashamed to Ask*. Chicago: University of Chicago Press.

McNeill, Daniel, and Paul Freiberger. 1993. *Fuzzy Logic*. New York, NY: Touchstone.

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.

———. 2017. "Qualitative Comparative Analysis and the Study of Non-State Actors." In *Researching Non-State Actors in International Security: Theory & Practice*, edited by Kruck, Andreas and Andrea Schneiker. Oxon: Routledge, 123-42.

North, Douglass C. 1984. "Government and the Cost of Exchange." *Journal of Economic History* 44 (2): 255-64.

Potter, Michael. 2004. *Set Theory and Its Philosophy: A Critical Introduction*. New York: Oxford University Press.

Putnam, Robert. 1988. "Diplomacy and Domestic Politics: The Logic of Two-Level Games." *International Organization* 42 (3): 427–60.

Quine, Willard Van Orman. 1969. *Set Theory and Its Logic*. Cambridge, MA: Harvard University Press.

———. 1982. *Methods of Logic*. 4th. Cambridge, MA: Harvard University Press.

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.

———. 2008a. "Measurement versus Calibration: A Set-Theoretic Approach." In *The Oxford Handbook of Political Methodology*, edited by Box-Steffensmeier, Janet M., Henry E. Brady and David Collier. Oxford: Oxford University Press, 174-98.

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

———. 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. 2017. *Intersectional Inequality: Race, Class, Test Scores, and Poverty*. Chicago: University of Chicago Press.

Ragin, Charles C., Susan E. Mayer, and Kriss A. Drass. 1984. "Assessing Discrimination: A Boolean Approach." *American Sociological Review* 49 (2): 221-34.

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.

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

Rubinson, Claude. 2019. "Presenting Qualitative Comparative Analysis: Notation, Tabular Layout, and Visualization." *Methodological Innovations* online first (July 18) (doi:10.1177/2059799119862110): 1-22.

Schimmelfennig, Frank. 2001. "The Community Trap: Liberal Norms, Rhetorical Action, and the Eastern Enlargement of the European Union." *International Organization* 55 (1): 47-80.

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

Shin, Sun-Joo. 1994. *The Logical Status of Diagrams*. New York: Cambridge University Press.

Smithson, Michael, and Jay Verkuilen. 2006. *Fuzzy Set Theory: Applications in the Social Sciences*. Thousand Oaks: Sage.

Tomassi, Paul. 1999. *Logic*. Oxon: Routledge.

United Nations. 2018. *Human Development Indices and Indicators*. New York, United Nations Development Programme.

Wagemann, Claudius. 2017. "Qualitative Comparative Analysis (QCA) and Set Theory." In *Oxford Research Encylopedia of Politics*, edited by Thompson, William R.: Oxford University Press, 1-36.

Whitesitt, J. Eldon. 2010. *Boolean Algebra and its Applications*. Mineola, NY: Dover.

Zadeh, Lotfi A. 1965. "Fuzzy Sets." *Information and Control* 8 (1): 338–53.