Chapter 2

Research Design

Mello, Patrick A. (2021) *Qualitative Comparative Analysis: An Introduction to Research Design and Application,* Washington, DC: Georgetown University Press, Chapter 2.

Research design ensures that the answers we provide are as valid as possible and are discovered as efficiently as possible.

Dimiter Toshkov[1]

Questions of research design are central to any scientific endeavor. To find appropriate solutions to social scientific problems, we first need to define our scope of inquiry, the concepts involved, and to lay out our theoretical expectations. Moreover, we need to specify which methods to use for the answers we seek to find.[2] Naturally, these general concerns equally apply for QCA studies – but there are additional points to consider. What kinds of *research questions* can be addressed with a set-theoretic approach? How should *theory* be formulated to gain most leverage from using QCA? How should *cases* and explanatory *conditions* be selected? Finally, can QCA be combined with other methods and how should such *multi-method* research be conducted? This chapter addresses these and some related questions in turn.

Despite the evident centrality of research design, this is often given less attention than matters of technical implementation and refinement at later stages in the research process. However, many of the challenges involved in conducting sensible research should arguably be addressed right at the outset of designing a study. For instance, a vexing issue for QCA is that the method conducts *static* comparisons. There is no formal place for time, sequencing, or process.[3]That said, nothing stops a researcher from incorporating these into the conceptualization of her conditions and outcome. Likewise, studies with a small or medium numbers of cases are at times criticized for overstating their conclusions (Lieberson 1991). Yet, a proper definition of scope conditions and case selection criteria can equally help to strengthen the inferences drawn from such studies as to designate their limitations. Finally, a sensible interpretation of QCA solution terms can pose a challenge in itself, especially if theory was not formulated in set-

theoretic terms. This can be averted by clearly spelling out observable implications and potential combinations of conditions derived from theory *before* the analytical part.

Present debates about research design typically take *Designing Social Inquiry* by Gary King, Robert Keohane, and Sydney Verba, or "KKV" (1994) as a starting point. In their influential book, KKV proposed a common inferential framework for scientific inference in quantitative and qualitative research. This ambitious goal naturally provoked criticism from those who liked to point out that, apart from causal inference, qualitative research aims for "[t]hick description and interpretation" (Caporaso 1995, 457), and that KKV "inappropriately view qualitative analysis almost exclusively through the optic of mainstream quantitative methods" (Brady and Collier 2004, xvi).[4] Be that as it may, one upside of the debate about the "qualitative-quantitative divide" is that it set in motion a process of renewed thinking about methods and methodology. The two editions of *Rethinking Social Inquiry* with their individual contributions (Brady and Collier 2004; 2010) and Charles Ragin's *Redesigning Social Inquiry* (2008) are landmarks in that regard. The volume edited by Henry Brady and David Collier provided a thorough critique of mainstream statistical approaches, while also building bridges and acknowledging common ground between qualitative and quantitative researchers. In turn, Ragin highlighted that despite his critical stance towards the "conventional quantitative template", he did not want his book to be understood as a critique of *Designing Social Inquiry*, but as a "middle path between quantitative and qualitative social research" (Ragin 2008, 1).

Research Questions

When working on a new scientific endeavor, one of the first tasks is to turn an idea of a research topic into a *researchable* project, which often goes in line with a research question. Among other suggestions, most scholars agree that research projects should strive to meet the two-fold criteria of (1) real-world relevance, and (2) making a scientific contribution to the scholarly literature in a given academic field (King et al. 1994; Schmitter 2008; Toshkov 2016).

But even once a topic has been defined, formulating the research question can be a tricky business. One pitfall is that questions can be too narrow or too specific, leaving little room for discovery or alternative explanations. For instance, based on a reading of relevant studies, you may ask whether there is a relationship between poverty and health issues. The advantage of such a research question is that it provides clear guidance for the analysis. There is no need to explore alternative factors. However, this setup may also be a straitjacket when you discover during your research that poverty is but one of several important factors that are related to health issues. Here, casting a wider net would have given your research more leverage. Relatedly, your research question may mention a specific case or several cases – again this may be fine because it provides focus, but it might also put you in a corner where it becomes difficult to account for relevant phenomena that only occur in some other cases. Certainly, these

concerns apply to all areas of the social sciences. In this chapter, we set aside such general issues and take it as given that we identified a research topic that meets the criterion of real-world relevance and that connects with scholarly debates in our field. Now our task is to turn this into a feasible comparative research project.[5]

But what are research questions that can be answered with QCA? And how are these typically phrased? Based on a two-fold differentiation by *focus* and *level*, four types of questions can be derived. Level differentiates between *case-specific* and *general* research questions. As the name implies, the former type asks about the circumstances of a specific case or several cases, whereas the latter asks a general question about relations between theoretical concepts without reference to specific cases. Focus separates *condition-centered* and *outcome-centered* research.[6] While the former seeks to understand the effect of a particular condition or combination of conditions (a *configuration* in QCA terminology), the latter aims to find the causes for an outcome, whether in one particular case or for a class of cases. Naturally, these aims can overlap in empirical research projects. With QCA, we may at times be equally interested in accounting for puzzling cases and identifying cross-case patterns. Nonetheless, the two-fold distinction helps to clarify the primary analytical aims of a project and to tailor the research design accordingly.

Another way to look at the differentiation between research questions is whether they arise primarily from puzzling empirics (case-specific level) or from theoretical considerations (general level). As for the first type, a researcher typically starts out from observing a case that presents an empirical puzzle. For example, why did British Prime Minister David Cameron initiate a referendum on leaving the European Union? Why was Italy affected so severely from the coronavirus pandemic? Why did social protests break out in Chile? All of these cases contain puzzling aspects. The questions can be framed around these outcomes, but they could also be centered on conditions such as Cameron's leadership characteristics, Italy's health care system, or social inequality in Chile. By contrast, the second type of research questions focuses on theoretical concepts or the relationship between several concepts. Here, the aim may be to explain the occurrence of a relevant outcome, or to examine whether a condition or a configuration leads to the outcome. For instance, rather than asking about the cases of Italy and Chile, we may be interested in generalizing across countries, to find out which conditions led to countries' being more affected by the pandemic or which led to the outbreak of social protests. Similarly, we could compare leaders' personality traits and place these in relation to certain outcomes.

Figure 2.1 summarizes the typology of research questions along the *focus* and *level* dimensions. The example is based on Spanish Prime Minister Zapatero's 2004 decision to withdraw from the Iraq War coalition. The first question remains case-specific, but it focuses on the impact that the condition of leadership change might have had on the withdrawal decision. By contrast, the second question highlights the outcome, asking for an explanation for this specific case.

Question three moves the inquiry to the general level, probing the effect of a specific condition, whereas question four asks about the general causes for the outcome.

Figure 2.1 Typology of Research Questions

		Focus: Condition-centered	Focus: Outcome-centered
Level: Case-specific		(1) How did leadership change affect Spain's withdrawal from the Iraq War coalition?	(2) Why did Prime Minister Zapatero announce Spain's withdrawal from the Iraq War coalition?
Level: General		(3) How does leadership change affect wartime coalitions?	(4) What are the causes of democratic coalition defection?

How does this look like in practice? Below are some examples of research questions from published QCA studies in business and management, international relations, political science, and sociology:

(1) "Under which conditions does spending on active labor market policies increase?"
(Vis 2011, 229)

(2) "Which policies promote the transition towards electric vehicles?"
(Held and Gerrits 2019, 13)

(3) "What explains the UN's selective response to humanitarian crises?"
(Binder 2015, 712)

(4) "How do the job attributes that comprise particular combinations of working conditions, or 'bundles,' interact to influence turnover decisions?"
(Nelson 2017, 24)

(5) "This raises the question if there are certain resources, or combinations thereof, that are necessary or sufficient conditions for civilians to employ control strategies successfully."
(Kuehn et al. 2017, 425)

The first question, from Barbara Vis' study on welfare state reform, asks "under which conditions" the outcome occurs. This implies that there is more than a single condition to account for the outcome and that specific combinations may be important. The second example, from Tobias Held and Lasse Gerrits' article on e-mobility similarly asks broadly about policies that "promote" the outcome (the adoption of electric vehicles), the implication being that several feasible policies exist. The study by Martin Binder, on the UN's response (and non-response) to humanitarian crises emphasizes an empirical observation ("The UN's selective response") and asks for an explanation for this puzzling finding. Research questions like this one can also be phrased generically along the lines of "why did X occur in [these cases], while it did not occur in [those cases]". The fourth example comes from Jennifer Nelson's article on the working conditions of schoolteachers. The study asks specifically about the interaction of "particular combinations of working conditions" that are expected to influence teachers' decisions on whether or not to leave a job. Finally, the article by David Kuehn and colleagues formulates an indirect research question about the occurrence of the outcome, framed in the language of necessary and sufficient conditions.

All of these examples are suitable ways to formulate research questions in QCA contexts because they resonate with the method's core assumptions. As we can see, all of them are variations of *general outcome-centered* research questions, though examples (4) and (5) also highlight the condition-side and causal complexity without referring to specific combinations. However, the prevalence of general outcome-centered questions should not be taken to rule out the other types (many examples can be found in the literature). Whenever there is a particularly prominent or puzzling case, a case-specific question may be a suitable starting point. It can also be complemented by a more general question about groups of cases. This can be a feasible strategy when a study aims to combine the QCA part with process tracing on a single case, for instance to identify a causal mechanism.[7]

Finally, you may ask whether it is indeed needed to formulate a research question, since many authors do *not* explicitly mention these in publications. Clearly, this is a matter of personal style and convention in a given academic field. Yet, from a research design perspective, I recommend narrowing down each project to a research question, because even if this will not be explicitly mentioned in the eventual publication, it can serve to structure the analytical process.

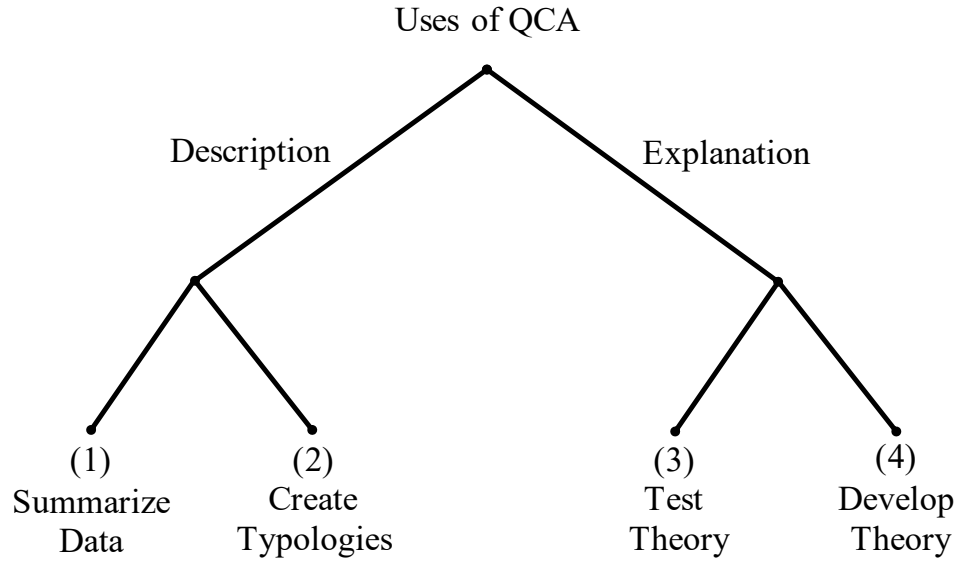
Uses of QCA

QCA can be used for a variety of different purposes, which may also differ by academic field and tradition. Yet four main uses of QCA can be distinguished, as summarized in Figure 2.2.[8] The core difference is whether the method is applied for the aim of *description* or *explanation*. When used for descriptive purposes, the method can serve (1) to *summarize and identify patterns in the data*. This can be helpful to see which cases share certain characteristics or to describe the

varieties under which an outcome occurs. For a handful of cases and, say, three conditions it may still be feasible to "eyeball" such patterns – but this swiftly becomes unwieldy as more cases and conditions are included. Here, the truth table routine and its systematic comparison of logically possible configurations present a clear advantage over less-structured alternatives. This way of using QCA also resonates with more *inductive* approaches that seek to explore the data without fully-specified theoretical expectations (Ragin and Rihoux 2004, 6). Another descriptive function that QCA can be used for is the (2) *creation of typologies*. This relates to established traditions in qualitative research (Collier et al. 2008; Elman 2005; George and Bennett 2005, Ch. 11) and it has been formalized as *fuzzy set ideal type analysis* (Kvist 1999; 2007). Chapter 8 looks into this QCA variant and an applied example.

When QCA is used for explanatory purposes, this is typically done for (3) *testing theory*. In fact, an overwhelming majority of QCA studies broadly fall into this category.[9] What these studies have in common is that they formulate theoretical expectations about the relationship between conditions and outcomes. As we will see shortly, there are different ways to do this. Finally, a variant of the explanatory use of QCA is for (4) *developing theory*. Here, the emphasis is not on conducting tests of existing theory, but to advance new theoretical arguments, often through the integration of various theoretical strands. That said, the categories of testing theory and developing theory are *not* mutually exclusive. Many applied settings contain elements of both, where established arguments from the literature are complemented with new theoretical elements to arrive at better explanatory accounts of the observed cases. Likewise, there are studies that would be rather located on the descriptive side, but which entail discussions of theoretical conjectures.

Figure 2.2 The Four Uses of QCA



Testing Hypotheses

When QCA is applied for the aim of *testing theory*, the most structured way of doing this is by *formulating hypotheses* and testing whether the theoretical expectations captured therein resonate with the empirical evidence (Amenta and Poulsen 1994, 29; Berg-Schlosser et al. 2009, 16). To be sure, this kind of usage rests on a broad conception of theory testing where a hypothesis is understood merely as a "tentative answer to a research problem, expressed in the form of a clearly stated relation" (Frankfort-Nachmias and Nachmias 2008, 56). This differs from more formalized ways of statistical hypothesis testing where the researcher formulates an experimental hypothesis and a null hypothesis and, ideally, proceeds to test these hypotheses on new data (Field et al. 2012).[10] Clearly, the latter kind of hypothesis testing would not be sensible in a QCA context, where a "dialogue between theory and evidence" is expected (Ragin 2014, xxi) and where good practice entails "carefully crafting the data" (Schneider and Wagemann 2012, 296). This has led some to caution against set-theoretic hypothesis testing altogether (Schneider et al. 2019, 7; Schneider and Wagemann 2012, 296).

At this stage, there is no consensus on the issue of hypothesis testing with QCA. While I share some of the concerns raised about it – and I will return to these at the end of this section, I believe that QCA applications *gain* inferential leverage from formulating their theoretical expectations in a clear-cut fashion. Certainly, this should not mimic deductive theory testing the way it is common practice in statistics. Yet, a clear explication of the *directionality* of each included condition and potential configurations enhances the interpretability of a QCA study because it allows an assessment of whether and how empirical results match theoretical expectations – and also what this might mean for some (yet) unobserved cases. That said, some requirements have to be met to do fruitful theory testing with QCA.

One prerequisite is that is that hypotheses must be framed in *set-theoretic terms*, meaning they use the language of necessary and sufficient conditions. Otherwise, there will be a mismatch between theory and methods. This can pose a problem when the research goal is to test established probabilistic hypotheses. While a large body of work in the social sciences rests on a (sometimes implicit) understanding of necessary and sufficient causation, many hypotheses in the literature remain framed in probabilistic language, requiring prior *translation* on the part of the researcher who seeks to employ such hypotheses in a QCA study (Goertz 2003a; 2003b).[11] For example, the following *probabilistic* hypothesis from international relations research on the "democratic peace" will be difficult to test with QCA:

> Hypothesis 1: "The more democratic a country, the more peaceful its external relations."

How can we confirm or disconfirm such a statement with QCA? We may expect that very democratic countries – consolidated democracies – are more peaceful than less democratic

countries. But we do not know how many exceptions we can tolerate to still consider our expectation confirmed. Assuming that we have plenty of peaceful democracies, how many instances of consolidated democracies engaged in war would be needed to disconfirm the above hypothesis? Moreover, another question is whether the relationship expressed in the hypothesis also holds for small increases in a country's "democraticness". This question may arise, for instance, when autocratic regime introduces democratic elements such as elections or a constitutional referendum. But we may doubt whether small-scale institutional changes also affect the country's peacefulness if the autocratic character of the regime remains untouched. In sum, probabilistic hypotheses require a *thorough clarification* of the underlying theoretical expectations to be used in set-theoretic contexts. On their own, probabilistic hypotheses are not helpful for QCA. Now consider this:

> Hypothesis 2: "Democratic political institutions are a sufficient condition for peaceful external relations."

We may doubt whether this hypothesis is empirically true – but clearly, testing it is a straightforward matter: the presence of the condition is expected to be sufficient for the outcome, which means that whenever we see the former, we should also observe the latter. Another point regarding hypotheses concerns *causal complexity* – the combination of conditions, alternate pathways toward an outcome, and causal asymmetry. While authors habitually point out conjunctural causation and equifinality as particular strengths of the QCA approach, these are infrequently incorporated on a theoretical level. How would a configurational hypothesis look like? For instance, we could reformulate the previous statement as a configurational proposition in set-theoretic terms, expecting two conditions to be jointly sufficient for the outcome:

> Hypothesis 3: "Democratic political institutions and societal norms of non-violent conflict management are jointly sufficient for peaceful external relations."

Finally, expectations about INUS conditions can be used for the formulation of hypotheses. In some sense, INUS conditions present a *hedged* way of formulating theoretical expectations, because they do not require an advance specification of the condition(s) that are expected to combine with the suspected INUS condition. Still, INUS conditions entail a *directional expectation* about the respective condition, requiring a judgment on whether it is the presence or rather the absence that should lead to the outcome. Here is an example from Barbara Vis' study on welfare state reform: [12]

> Hypothesis 4: "Rightist partisanship is an INUS condition for unpopular reform." (Vis 2010, 138)

This leaves the question of whether formulating hypotheses can be reconciled with the back-and-forth between data and evidence that is characteristic of QCA. Surely, it would make no sense to first conduct a QCA and then formulate the resulting complex paths as "hypotheses" in the theory section. Hence, it often is sensible to begin with *modest* directional expectations (for instance about INUS conditions) and to justify these expectations in reference to the existing literature on a topic. When this is done, the expected relationships can be examined in the set-theoretic analysis. This can either lead to theory confirmation or the discovery of combinations of conditions that help to re-formulate and modify the existing theory. Indeed, as Ragin (2000, 58) states, if a modification is "part of an attempt to learn more about the world […] and not part of explicit program of theory testing, it is completely reasonable."

Case Selection

An integral part of qualitative research design is the selection of cases. When thinking about case selection, several questions must be addressed: What are my cases? How can I distinguish positive from negative cases? How many cases shall I include? Should the cases be sampled from a population or comprise the entire universe of cases?[13] The following sections address these questions in turn.[14]

What Is a Case?

Simply put, a case is a single row in the QCA data sheet. More substantively, a case may be defined as the *unit of analysis* or a "spatially delimited phenomenon (a unit) observed at a single point in time or over some period of time" (Gerring 2007, 19). This means that a case can be anything we are interested in, including a country, a government, an organization, a company, a village, a piece of legislation, or an individual. What is more, we can take into account *time* to separate our case (the unit) into several different sub-units, if that helps our purposes. For instance, by looking at the same country over different years, or examining an organization at different stages of its institutional development. Because of the implications for everything that follows, defining cases or units of observation should be among the first tasks in designing a QCA study.

That said, *revisiting* the case definition, and asking the essential question *what is this a case of?* (Ragin and Becker 1992), can also be a powerful tool to improve a study if problems occur during the analysis or the interpretation of the findings. For example, we may start out with a comparative analysis of the 27 EU member states, but later realize that we should differentiate between individual government cabinets, because of policy changes that occurred from one prime minister to another. Hence, we could revise our research design to include multiple different government cabinets per country, over a specified period of time. This would increase the total number of cases and it would require new data gathering, calibration, and a reanalysis

of our data. The benefit of such a procedure is that it would make our analysis more fine-grained and it would allow for the inclusion of additional conditions related to the adapted unit level, for instance on the political partisanship of a government or the leadership traits of prime ministers.

Populations, Scope Conditions, and Samples

We can distinguish five ways of how cases can be selected for QCA. The first option is to select all of the cases from a (1) *given population*. For example, we could include all 30 companies listed in the Dow Jones Industrial Average stock market index, the 27 member states of the European Union, or the 22 members of Boris Johnson's second cabinet. This approach means that once we have decided on a given population, the case selection itself should be straightforward and unproblematic. Given populations have the benefit of *face validity* because they are not subject to a researcher's judgment about which cases to include or not to include. However, the caveat is that even supposedly "given" populations tend to change over time, and hence we must specify the time period we seek to cover in our analysis. The Dow Jones index from 1982 certainly looks different than the one from 2017 and the number of EU member states has changed considerably over the past two decades. Another downside of given populations is that these may contain *irrelevant cases*, to which we turn in the next section.

The second approach is similar, but further introduces (2) *scope conditions* to limit the selection of cases. Scope conditions are used to bring the range of cases in line with theory, as most social science theories are not meant to be universal but rather apply to a limited area of application.[15] This resonates with the conception of middle-range theory (Merton 1958). For instance, there may be theoretical reasons why we would want to consider only companies that have been listed in the Dow Jones for the past two decades. Or why we prefer to restrict our analysis to EU member states with at least five million inhabitants. Finally, we might want to include only those members of Johnson's cabinet who supported the "Leave" campaign during the Brexit referendum in 2016. To take another example, studies that work with the Polity IV data on political regimes (Marshall et al. 2019) typically introduce a threshold for the combined autocracy-democracy scale (which runs from -10 to +10), above which countries are included in the case selection. This may rest on the assumption that the theory is only valid for consolidated democracies, rather than defective democracies, or hybrid regimes. Here, a threshold of 7, for instance, may serve as a scope condition for the case selection where just those countries are included that meet or exceed this threshold. Lastly, it should be noted that scope conditions can be combined with any of the other approaches to case selection, as in limiting a given population through a scope condition (Ragin 2000, Ch. 2).

The third approach is to (3) *purposefully select* based on cases' value on the outcome. In the quantitative tradition, this is often seen as a violation of a core principle ("never select on the

dependent variable"), but for qualitative research this is a vital tool to focus on relevant cases.[16] For instance, imagine you want to study why some companies are more successful than others in implementing family-friendly working conditions. Your resources allow you to analyze not more than 20 companies, also because you plan to conduct interviews at each of these, but you have a population of *several hundred* companies that meet your criteria. Now the fourth approach would be taking a (4) *random sample* of 20 companies, but it may be that this sample of randomly drawn cases will contain none or only a handful of positive cases with successful implementation of family-friendly working conditions. It may also be that the random sample contains not a single typical case where the outcome and an expected causal mechanism are both present. This is why random sampling "is not necessarily a wise technique to use" for qualitative research with small to medium numbers of cases (King et al. 1994, 125).

A more sensible approach for the chosen example would be to purposefully select 10 positive and 10 negative cases to compare their configurations of conditions and to identify difference-makers between the two groups. As Ragin (2000, 59) notes, "researchers also may constitute a population of negative cases to compare with the positive cases." In our example, this would mean that we begin with a number of cases that show the phenomenon of interest and match this number with an equal number of cases that do not. Finally, a combined approach would be using a (5) *stratified sample* of cases. This means that the population is first divided into a number of meaningful subsets and from each of these subsets a random sample of cases is drawn in the second stage. Thereby the population is effectively reduced, while retaining representativeness.[17] This approach works best with large-*N* data and it requires some prior information on which to divide the population into subsets of cases.

Most QCA studies base their case selection on *given populations*, *scope conditions*, or *purposeful selection*. These approaches also resonate with qualitative research methods. That said, I have further included *random sampling* and *stratified sampling* as rather quantitative approaches to case selection, and thus less often seen in QCA studies, because there are research settings where the advantages of sampling may outweigh the downsides to case selection. This may be, for instance, in sociological, medical, or business applications where large-*N* data on individuals, patients, or companies is worked with. Table 2.1 summarizes the five approaches to case selection.

Table 2.1 Case Selection Approaches

Approach	Characteristics	Strengths	Limitations
Given population	Predefined, given population of cases	High face validity	May contain irrelevant cases
Scope condition	Limits a theory's assumed scope of validity	Resonates with middle-range theorizing	Requires careful justification; may introduce bias
Purposeful selection	Selection on the outcome (and/or the non-outcome)	Selection is guided by relevancy of cases	Requires careful justification; may introduce bias
Random sample	Random sampling from a population of cases	Representative sample, generalizable	May exclude important cases; requires large-N
Stratified sample	Random sampling among subsets of the population	Adequate representation of cases	Requires large-N and prior information on which to divide into subsets

Positive, Negative, and Irrelevant Cases

When designing a study, we typically begin by thinking about *positive cases*. These are cases that have sparked our interest in a topic because they represent a new empirical phenomenon or because they contradict or resonate with our theories. At the early stage of research design, effort should be placed on conceptualizing the positive outcome and distinguishing it from neighboring concepts. Often, this entails a *back-and-forth* between theory and cases as we discover that our outcome, as we conceptualized it, does not do justice to important cases (Ragin 2000, Ch. 2). However, apart from the positive cases, for our research design we should also think about negative and irrelevant cases.

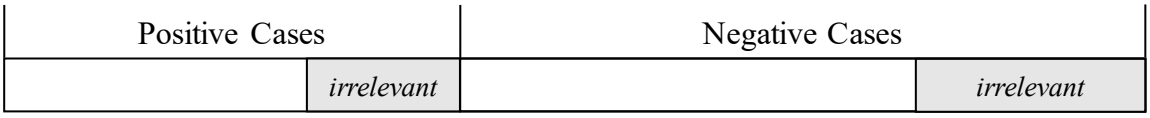
Upon first thought, *negative cases* may appear uncomplicated: these are cases that do *not* show the outcome. But for many concepts it can be challenging to pin down what its non-occurrence actually means. To take the example of Theda Skocpol's work (1979) on social revolutions, we may ask what it means to have a "non-social revolution" (Goertz and Mahoney 2006, 177). Similarly, what is the negation of "war involvement"? Is that involvement that is non-military (as in political or humanitarian involvement), or does it refer to the absence of any kind of involvement? Often, the negation of a concept can entail a variety of meanings. For instance, "non-democracy" might include monarchies, authoritarian states, and dictatorships. And the negation of a successful policy response to the coronavirus might include ill-suited measures, belated responses, and non-responses. The bottom line is that QCA researchers are well advised

to clarify their intended meaning of the non-outcome, and more generally, to specify which criteria a case needs to fulfil in order to be considered inside or outside a set (see also Chapter 5).

Another related category are *irrelevant cases*. As Ragin (2014, xxvi) notes, given populations often comprise cases that are deemed "irrelevant", which can mean that cases are included that neither show the outcome nor hold meaningful values on the explanatory conditions. For example, let us assume we want to study the conditions under which foreign aid is provided in humanitarian emergencies, based on the given population of the 193 UN member states. However, using this given population means that countries are included in the sample that may lack the economic and financial capacity to provide foreign aid in the first place. Hence, including such cases in our study would inflate the number of negative cases and may thus lead to flawed inferences based on conditions shared by the irrelevant negative cases. Irrelevant cases *partially* overlap with the set of "impossible cases", to use Gary Goertz's and James Mahoney's term (2006). The lesson for case selection is that one should only include those cases where the outcome *could possibly happen* given what we know about the case. This implies that cases where the outcome is deemed impossible should be treated as "uninformative and hence irrelevant observations" (Goertz and Mahoney 2006, 179).

Figure 2.3 shows the relationship between positive, negative, and irrelevant cases. For many phenomena that social scientists are interested in, the size of the set of negative cases exceeds the set of positive cases, as indicated in the illustration. Importantly, both positive and negative cases may entail irrelevant cases that should be excluded. These may be cases for which it would be impossible to show the outcome. Likewise, there might be cases that show the outcome, but given what we know about them, it would have been impossible for them to *not* show the outcome. For example, suppose you study why some countries abstained from implementing a certain EU policy at a given point in time. Now there might be a country that you would treat as a "positive" case because it did not implement the policy. However, closer inspection might show that the country experienced a constitutional crisis at the time and there was no functioning government to implement the policy. Hence, it might not be sensible to include this as a "positive" case of your outcome. Apart from cases where the outcome or the non-outcome are impossible, the set of irrelevant cases can further entail those that are deemed substantively unimportant and which should thus be excluded from case selection. For example, suppose we wanted to study how various political factors impact upon parliamentary debates in consolidated democracies (the number of parties, their ideological positions, and so forth). Now there might be a country that fulfills our criteria for case selection, as it meets a certain democracy threshold, but where parliament was dissolved during the observed timeframe, owing to a constitutional crisis. Hence, we may consider this case irrelevant for substantive reasons. Ultimately, whether or not a case can be considered substantively unimportant depends on the research aims of a given study.

Figure 2.3 Positive, Negative, and Irrelevant Cases



How Many Cases Are Needed?

There is no golden rule as to how many cases *must* be included in a QCA study. Nor is there an upper limit for the inclusion of cases. That said, QCA works well with medium numbers and this shows in the fact that in many academic fields the majority of published articles draw on a range of 15 to 30 cases, as reflected in the survey results discussed at the end of this chapter.[18]

More important than the absolute number of cases are two related indicators: (1) *the ratio between cases and conditions*, and (2) *the empirical distribution of cases across truth table rows*. The first indicator is relevant because the number of conditions determines the size of the truth table. With each condition that is added, there will be more logically possible combinations that cases can fall into. This means that, if this is repeatedly done, eventually there will be one row for each case (and many empty rows). The side effect is that *consistency* – a concept to be discussed in Chapter 6 – increases when conditions are added, but the resulting solutions may or may not be meaningful, because even random data may generate patterns of sufficient configurations when there are too few cases per condition (Marx and Duşa 2011).[19] Hence, since the truth table grows exponentially with each condition that one adds to a study, one should simultaneously increase the number of cases. This does not mean that the number of cases has to match the number of possible configurations, but we should keep in mind the size of the truth table when thinking about how many cases we want to include in our study.

As an orientation mark, I suggest using a ratio of *at least* four cases per condition, and higher ratios if five or more conditions are included because of the exponential growth of the truth table. This means that a standard QCA application with 4 or 5 conditions should entail at *minimum* 16 or 25 cases, respectively (see Table 2.2). This recommendation broadly resonates with earlier efforts to formulate benchmarks for the inclusion of cases, as suggested by Axel Marx and Adrian Duşa (2011, 114), who ran simulations with crisp-set data (see also Marx 2006).

The second indicator, the empirical distribution of cases across truth table rows, reveals the *empirical scope* of the research design. This is often overlooked because discussions tend to focus on the absolute number of cases. Yet, this may even be a more critical factor to consider, as studies can differ greatly in the extent to which the empirical cases distribute among the logically

possible combinations of conditions. Imagine you may have a research design with 16 cases and three conditions where the cases evenly fill out all of the eight logically possible rows with two cases per row. This would be a situation with a *complete* truth table without any logical remainders (and hence the solution terms that you would eventually derive from this would be identical, as will be discussed in Chapter 7). In another scenario, you might also have 16 cases and three conditions, but the cases cluster together in only three rows, leaving five logical remainders. This means that although you may have plenty of cases, there is little variation between them. This prompts questions about whether further cases should be added to reduce limited diversity and whether the selection and conceptualization of the conditions should be changed to achieve a more even distribution across the truth table rows.[20]

Now one may ask why the number of empty rows matters – why is it that we cannot simply work with the cases at hand, regardless of limited diversity and logical remainder rows? To be sure, the technical routine of QCA works even under extreme scenarios where, let's say, conditions outnumber cases. However, besides from the inferential problems mentioned above, looking at the *empirical scope* of a study can help us to identify problems at an early stage. When cases cluster in but a few rows, leaving many logical remainders, then this suggests that there might be a *mismatch* between the explanatory model and the case selection. After all, there are simply no cases for many implied combinations of conditions. The apparent mismatch between the empirical variety and the logically possible variety, as in the above example, may result from the criteria that were applied for the selection of cases. If we had included a larger number of cases, then these might have filled more truth table rows. The mismatch might also result from our conceptualization of conditions. If conditions had drawn on different indicators, then the cases might have distributed more evenly. Even simpler, it might be that some of our conditions, while theoretically feasible, do not resonate with the empirics of our research design. Hence, revisiting and possibly replacing some conditions might be warranted.

Table 2.2 illustrates the relationship between the number of conditions and truth table rows, as well as the recommended minimum numbers of cases. The latter is based on a suggested ratio of cases per condition, starting with a minimum value of four cases per condition and increasing from five conditions upward. These recommended ratios are meant to provide *general advice* on how many cases should be included for a certain number of conditions – as this is one of most frequent questions asked by new users of QCA.

The recommended ratio comes with the *caveat* that such thresholds should never be applied mechanically, nor can they replace individual judgment. In some research settings, one simply has to make do with a small number of cases. For instance, this can be because of the small size of a given population or for reasons of missing data. However, in both scenarios a researcher could still decide to focus on a smaller set of conditions and thereby limit the size of the truth table. From this perspective, there are always two possible adjustments to avoid a mismatch

between conditions and cases: one can decide to either *increase* the number of cases or to *decrease* the number of conditions included.

Table 2.2 Cases, Conditions, and Truth Table Rows

Number of Conditions	Truth Table Rows	*Suggested* Minimum Number of Cases	Ratio of Cases per Condition
2	4	8	4
3	8	12	4
4	16	16	4
5	32	25	5
6	64	36	6
7	128	42	6
8	256	56	7

Condition Selection

The selection of conditions is central to designing effective research with QCA. This selection compels us to balance opposing tendencies: on the one hand, we often want to include as many conditions as possible, to provide a full account of the phenomenon we seek to explain. On the other hand, each added condition makes our analysis more complex and thus more difficult to interpret. Adding conditions increases the number of possible configurations (shown in the size of the truth table) and with it the scope of *limited diversity*, as there will be fewer empirical cases for each combination. Albert Einstein is reported to have said "everything should be made as simple as possible, but not simpler" (Sessions 1950). In line with this adage, as QCA researchers, we may want to keep the number of conditions small, while allowing for enough complexity to investigate various configurations of relevant conditions.

Apart from balancing these two tendencies, QCA studies should always provide a careful *justification* for their selection of conditions. This serves to address a common response, as reviewers and audiences often rightfully ask why certain conditions were *not* included in a study (and how the results would have looked like if these had been included). To address such concerns, QCA studies should strive to (1) include all important conditions, or as many of these as feasible, and (2) provide a thorough justification why these and not others were selected.

Plausible reasons why certain factors were not considered as conditions can be manifold: it can be due to a lack of available data (when there is no accessible information on some of the cases), because a considered factor does not vary across the observed cases (as when all cases hold similar or identical values on a condition), or owing to boundaries of the theoretical framework (which may specify certain kinds of factors but not others).

To reduce the number of conditions, we can also apply *scope conditions* by limiting the analysis to a certain domain or level. This is similar to the usage of scope conditions in case selection, as discussed previously. For example, we may restrict our theoretical framework to the formal constitutional structures of democracies, hence excluding factors such as political culture or societal norms, both of which would be outside that scope. In another setting, we may restrict our analysis to a variety of personality traits among a group of entrepreneurs. This would rule out external factors, such as the socio-economic environment that the respective businesses operate in (something that we should address in our case selection).

The number of conditions can also be reduced when there are several similar or related factors and including one of them suffices for our purposes. For instance, if our study already includes a condition "green party support" based on electoral votes for a green party, then we may not need another similar condition that reflects public preferences about green policy issues. If we do not want to lose information, we can also aggregate several related factors into index conditions or "macro-conditions" (Ragin and Fiss 2017, 75). This can be an effective way to take into account the various dimensions of higher-level concepts, such as "democracy", or "social equality". How this is done in practice is a matter of calibration, which we discuss in Chapter 5. That said, regardless of which and how many conditions we eventually include in our study, there should be a discussion of *why* these were selected and possibly, why certain other conditions were *not* considered.[21]

What is important to note in this context is that in QCA there are *no control variables*. This is another difference in relation to statistical methods, where control variables are frequently used to test for spuriousness (Frankfort-Nachmias and Nachmias 2008, 50). In statistics, this is done to ensure that an observed relation between an independent and a dependent variable is not caused by some third variable (the control variable). By contrast, all QCA conditions have the *same status* and are treated the same way throughout the analysis.[22] Therefore, it would make no sense for a QCA study to label some conditions as "controls" because all of them are part of the same analysis.

How to begin with the selection of conditions? Methodologists have suggested various ways of selecting and reducing the number of conditions used in QCA studies (Amenta and Poulsen 1994; Berg-Schlosser and De Meur 2009; Kahwati and Kane 2020; Schneider and Wagemann 2006). While there are no firm rules on condition selection and the context of a given study should always be kept in mind, we can still derive some guidelines. First, as mentioned earlier, the *number of conditions should be kept as small as feasible*. The strengths of QCA show when the number of conditions is in the small to medium range (say three to five conditions). Working in that range allows the researcher to properly theorize the conditions and their expected relation to the outcome, and to discuss potential interaction between conditions. A small number of conditions is also helpful to focus on key aspects and to highlight individual paths in QCA solution terms. Yet, this should not be taken to imply that more conditions cannot be

managed – as there are examples of successful studies with more than five conditions – but larger numbers of conditions do pose particular challenges, and this should be kept in mind when designing a study.[23]

Second, *conditions should be conceptualized in a way that resonates with causal complexity*. The identification of causal complexity is a particular strength of QCA. Hence, to make the most of QCA, studies should conceptualize their conditions in a way that builds on this strength o the method. This means that interacting elements and causal mechanisms should be conceptualized as separate conditions that can be analyzed empirically. The conditions should also allow for conjunctural causation and the existence of multiple pathways or recipes. Finally, *conditions should be conceptualized in a straightforward manner and not be too narrow*. The former ensures that it remains possible to interpret what it means when a case holds membership in a given condition, whereas the latter cautions that conditions should have broad applicability across cases. Because we can only accommodate a limited number of conditions, these should not be used to cover idiosyncrasies of specific cases (for instance, the particularities of an electoral system that can only be found in three out of thirty cases).

Empirically speaking, we can distinguish five approaches to the selection of conditions. These are summarized in Table 2.3. The *single-model* approach applies a stable set of conditions throughout the analysis of the outcome and the non-outcome. The conditions themselves are located at the same level or part of the same theory or theoretical framework. This may be the most commonly used approach among QCA studies. Examples include the articles by Boogaerts (2018), Pullum (2016), and Vis (2011). For instance, the study by Andreas Boogaerts (2018) aims to explain the European Union's use of sanctions during the Arab Spring. Since the study works with a relatively small number of only 13 cases, it constructs two *macro-conditions* that combine several indicators on the violent suppression of protests and material and security interests, in addition to two regular conditions, for a total of four conditions (Boogaerts 2018, 414). An advantage of the single model approach lies in its simplicity: it is easy for readers to grasp which and how many conditions were used, and the results can be documented in a straightforward fashion.

The *integrated model* approach differs from the single model approach only in the sense that conditions are explicitly drawn from different theories and integrated into a single stable set of conditions. For example, for studies in international relations, some of the conditions may be derived from theories that are based on the level of the international system, whereas other conditions are rooted in the level of domestic politics (e.g. Haesebrouck 2017; Mello 2020). Along those lines, Tim Haesebrouck (2017) combines international factors with domestic constraints to provide an integrated model of burden sharing among NATO member states. The integrated approach works well with QCA because it often entails concrete expectations about the interaction between certain conditions, which can also be visualized.[24] The challenge

is to provide a persuasive justification for the selection of conditions because the candidate pool of potential conditions tends to be larger than for studies with a single theoretical framework.

The other three approaches all entail several analytical steps. As the name implies, the *exploratory models* approach is based on several models of conditions, which are tested for their inferential value in terms of explaining the outcome. Typically, this approach is used when there is a broad but inconclusive literature on a topic (Ahn and Lee 2012; Avdagic 2010; Pinfari 2011). For instance, in her study of social pacts between governments, employers, and unions, Sabina Avdagic sets out "to explore multiple models, each containing a modest number of conditions" (Avdagic 2010, 643). Altogether, Avdagic derives 11 conditions from the political economy literature, but these are narrowed down to five conditions over the course of the analysis. The strength of this approach lies in its inclusiveness, as it covers all or nearly all potential conditions that are mentioned in the literature. But this comes at a cost, because the analytical steps are multiplied by the number of models, which poses challenges for the substantive interpretation and the documentation of the results.

The *competing models* approach works in similar fashion, but here the models are drawn from rival theories or schools of thought and placed against each other in sequential analyses (Hörisch 2013; Lilliefeldt 2012; Maat 2011). For instance, Felix Hörisch (2013) derives conditions from the competing theories of partisan politics and varieties of capitalism and Emelie Lilliefeldt (2012) draws on theories of social democracy and Protestantism, among others. Eelco Van der Maat (2011) pits different models based on theories of international relations against each other, including realism, neoliberal institutionalism, and domestic-level approaches. In his study, van der Maat conducts consecutive analyses of these models and also includes a combination of the latter two models. In his empirical analysis, the realist model emerges as the framework with the greatest explanatory power (highest consistency and coverage values). The competing models approach benefits from its comprehensiveness and its potential to compare the empirical resonance of different theoretical accounts. Apart from the increased complexity, which it shares with the exploratory approach, an additional challenge lies in developing models that properly reflect the respective theories in an unbiased manner.

Finally, the *two-step approach*, developed by Carsten Schneider and Claudius Wagemann (2006) differentiates between remote and proximate conditions and runs sequential analyses to identify contexts under which certain configurations lead to the outcome. Using the two-step approach, Maria Brockhaus and colleagues (2017) analyze the implementation of environmental protection policies by distinguishing between institutional settings (conceived as remote conditions) and the policy arena (proximate conditions).[25] Similarly, Julian Kirchherr, Mats-Philip Ahrenshop, and Katrina Charles (2019) examine large dam projects and differentiate between structural factors that are largely stable (remote conditions) for a single dam project and those that vary, like the project cycle and political stakeholders (proximate conditions). The two-step approach resonates with the way in which explanations are often

formed in the social sciences. Typically, there are context factors that may positively impact upon the outcome (or which may be necessary), but these may not in themselves provide sufficient explanations. Hence, there is the need to combine these with factors that are closer to the phenomenon that ought to be explained. As with some of the previous approaches, the downside to this is increased complexity due to the sequential analytical steps (see Chapter 8 for an illustrated example of the two-step approach).

What is the most suitable approach to selecting conditions? Evidently, there is no silver bullet for arriving at a sensible number of conditions. A core question is whether to work with one model of conditions or to include several models. The first two approaches apply the same model throughout, whereas the latter three approaches use sequential analyses with several models. Clearly, using multiple models allows for more conditions to be included in the analysis. This can be helpful to test rival theories or to explore a broader number of theoretical conjectures based on comprehensive literature reviews. The downside of a larger number of conditions and multiple models is that the analysis swiftly becomes unwieldy. For publications, particularly journal articles, this can pose a problem because there will be less space to discuss individual conditions and to explore the results. Ultimately, the decision of whether to use a single model or several models should be made in line with the research aims and the theoretical basis of a given study. If one is unsure about the inclusion of certain conditions, I suggest starting with a handful of the most important conditions and slowly increasing the complexity until a satisfactory explanatory model is found. Building up a study in an incremental and iterative manner, through a back-and-forth between theory and evidence, helps to identify patterns among the cases and relationships between the conditions, which can be further refined by adding conditions, cases, and successive analytical steps.

Table 2.3 Approaches for the Selection of Explanatory Conditions

Approach	Characteristics	Examples
Single-model approach	Stable set of conditions throughout the analysis	Boogaerts (2018), Pullum (2016), Vis (2011)
Integrated model	Stable set of conditions based on theoretical integration	Haesebrouck (2017), Mello (2012)
Exploratory models	Multiple models of conditions, used for explorative purposes	Avdagic (2010), Ahn/Lee (2012), Pinfari (2011)
Competing models	Multiple models of conditions, used to test rival theories	Hörisch (2013), Lilliefeldt (2012), Maat (2011)
Two-step approach	Remote and proximate conditions, sequential analyses	Brockhaus et al. (2017), Kirchherr et al. (2019)

Multi-Method Research Designs

In many areas of the social sciences, multi-method research designs have become the *gold standard* for empirical studies. Publications and collaborative research projects show that there has been a virtual multi-method "boom" during the past decade as *nested analysis*, *mixed method* and *multi-method* research strategies have gained considerable popularity (Beach 2020; Goertz 2017; Lieberman 2005; 2015; Seawright 2016).[26] This also applies to QCA, where a range of frameworks have been proposed on how to conduct multi-method research in set-theoretic contexts (Beach and Rohlfing 2018; Kahwati and Kane 2020; Meegdenburg and Mello forthcoming; Pattyn et al. 2020; Rihoux et al. 2021; Rohlfing and Schneider 2018; Schneider and Rohlfing 2013; 2019). Arguably, these trends are the result of both a broader recognition of the inherent limitations of social science methods as much as they are an acknowledgment of causal pluralism, which can be tapped into by using different inferential approaches (Anjum and Mumford 2018; Illari and Russo 2014).

But what are the advantages of multi-method research and how can it be implemented with QCA? Before addressing this question, it is vital to note that QCA, in itself, can be considered a multi-method approach, since it is based equally on the *qualitative* study of cases and *quantitative* analytical procedures (Rihoux et al. 2009). This requires intimate case knowledge, even though the level of detail will differ with the research aims and the number of cases in a given study. Yet, as Benoît Rihoux and Bojana Lobe rightly highlight, "at virtually every step of the QCA procedure, there is a dialogue with the individual cases" (2009, 229). Moreover, many set-theoretic studies draw on both qualitative and quantitative forms of data, as one of the characteristics of a mixed methods approach (Kahwati and Kane 2020, 12). As such, the outcome to be explained may rest on qualitative information that is gathered from interviews, official documents, media reports, and the like, whereas explanatory conditions may be based on pre-existing quantitative data, such as economic or social indicators, or official statistics and other sources of numerical information. From this perspective, it may appear misguided to demand that QCA must *always* be combined with other methods, without taking into account the specific research context a study situates itself in.[27]

That said, many scholars agree that the combination of several research methods holds potential because the inferential strengths of one method can be combined with those of another, and the blind spots of each may be overcome with a sound multi-method research design (Beach and Pedersen 2013; Blatter and Haverland 2012; Gerring 2012; Goertz 2017; Seawright 2016).[28] As will be explored in Chapter 4 with its emphasis on causal pluralism, different methods are suitable for different kinds of causal explanations. For instance, while an intensive case study may provide us with a good grasp of a causal process at play in a single case, it would not necessarily allow us to engage in generalization to a larger number of cases. With QCA, we may be able to identify a cross-case pattern, but this might not suffice to uncover a causal

mechanism. For that, we would have to examine whether the observable implications of our theory can be identified during the actual process as it evolves over time in a given case.

To be sure, there are different understandings of what multi-method research entails. One conception that is not considered here is the combination of large-*N* statistical analysis with case studies (Lieberman 2005; 2015; Seawright 2016) and, specifically, the combination of statistics and QCA (Meuer and Rupietta 2017).[29] In the context of set-theoretic methods, most work has focused on the combination of QCA as a method for *cross-case* analysis and process tracing as a method for *within-case* analysis, and, specifically, for the identification of causal mechanisms (Beach 2018; Beach et al. 2019; Beach and Rohlfing 2018; Meegdenburg and Mello forthcoming; Pattyn et al. 2020; Rihoux et al. 2021; Rohlfing and Schneider 2018; Schneider and Rohlfing 2013; 2019). While a detailed discussion of the suggested frameworks is beyond the scope of this section, four approaches can be distinguished, all of which focus on combining QCA and process tracing.

The first question when engaging in multi-method research with QCA and process tracing is whether to conduct the within-case analysis *before* or *after* the QCA part. Before QCA, the within-case analysis can serve, first, to *explore* potential conditions and to gain a better understanding of a prominent case. For this purpose, we would usually select a case that shows the outcome of interest, but we may not yet know how this case relates to the larger population of cases. Such an exploratory use of case studies can serve to identify relevant conditions and to build hypotheses (Rohlfing 2012, 11). In an ideal scenario, the exploration may even yield a mechanistic explanation for the outcome (Beach and Pedersen 2019, 9). A more formalized way of doing case studies before QCA is selecting a *typical case* based on a larger cross-case relationship. Clearly, this requires some kind of prior analysis to establish whether or not a case can be deemed typical, or representative of a population of cases (Gerring 2007, 91-97). In a QCA context, it may be that a topic has been well-researched, so that it is possible to identify a typical case based on existing studies in the field. Process tracing on this case would aim to identify a within-case relationship between a condition or configuration and the outcome, or to test a hypothesis about the presence of a condition or combination of conditions (Beach and Rohlfing 2018, 12). Both of these uses of case studies may require some iterations on additional cases if the results are not conclusive, before one can proceed with the cross-case QCA part.

When intensive case studies are conducted after QCA, they can either serve to confirm theory or to modify and further develop the theoretical framework (Schneider and Rohlfing 2013). For this usage, we can select a *typical case* from the QCA results, which is a case that holds membership in the solution term and the outcome. More precisely, we should select a *pathway case* (Gerring 2007, 122). For the QCA context, a pathway case would be a case that is uniquely covered by one of the solution paths. The alternative is to select a *deviant case*, defined as a case that holds membership in the solution, but which does not show the expected outcome.[30] Depending on the results, these case analyses may provide confirmatory evidence of causal

mechanisms, but they may also identify shortcomings in the theoretical account as in overlooked factors or unexpected effects, which may lead to a modification of the theory and a renewed cross-case case analysis with QCA. The general contours of the four basic approaches to combining QCA and process tracing are summarized in Table 2.4.

Table 2.4 Combining QCA and Process Tracing

Research Phase	Process Tracing	Aims	Consequences
Before QCA	Exploratory process tracing	Identify relevant conditions	Follow-up with QCA (*cross-case analysis*)
	Process tracing on a typical case	Identify within-case relationship	
After QCA	Process tracing on a pathway case	Identify causal mechanism	Confirm theory
	Process tracing on a deviant case	Identify room for theory improvement	Modify theory

Box 2.1 Multi-Method QCA: Environmental Peacemaking (Ide 2018)

Environmental Cooperation and Peacemaking Between States: A Multimethod Research Design Combining Statistical Analysis, QCA, and Case Studies

By Tobias Ide (*School of Geography, University of Melbourne*)

Researchers have long speculated that environmental cooperation can facilitate peacemaking between states ("environmental peacemaking"), but little cross-case evidence on this hypothesis existed. In order to fill this gap, I employed QCA in the context of a multi-method research design (Ide 2018). Specifically, I aimed to study the links between cooperative environmental agreements (which might help to build trust or create interdependence) and the termination of international rivalries (by means of reconciliation).

In a first step, I employed odds ratio tests of all possible cases in the sample. These indicated a weak, yet significant link between the conclusion of an environmental agreement and rivalry termination. Based on the odds ratios and the existing literature, I concluded that environmental peacemaking is highly context dependent. It is here where QCA kicked in. I selected the 20 cases in the sample in which rival states signed a cooperative environmental agreement and identified the conditions distinguishing the 6 case of rivalry termination (in the five years after the agreement) from the 14 cases without this outcome. Results show that the

combination of high environmental attention, political stability and conservation cooperation is sufficient for environmental peacemaking.

In my view, the multi-method research design was key to convincing the reviewers. The total sample of cases (dyadic rivalries in the 1946-2010 period) would have been too large for the QCA. And starting the study outright with the 20 cases was infeasible as the key explanatory variable of the study (cooperative environmental agreements) would have shown no variation. So I decided to use statistical techniques in order to answer "whether" questions and QCA to answer "when" (or "in which context") questions.

Subsequent desk-based qualitative studies of the 6 positive cases significantly refined the QCA results. For example, they illustrated the causal links indicated by the QCA, but also qualified that environmental agreements catalyze, but do not initiate processes of reconciliation in rivalries (hence addressing "how" questions). Further on, the case studies identified one outcome not explained by the QCA as a false positive and uncovered an additional relevant causal condition: pre-existing, informal environmental cooperation.

So QCA shows considerable potential for application in multi-method research designs. But this is not always easy to do. Reviewers can be (and in my case in fact were) skeptical whether, for instance, it would be better to do one in-depth rather than three supposedly superficial analyses. Extensive references, a concise writing style and the preparation of informative online appendices can be helpful here. One should also keep in mind that reviewers are often more familiar with statistical and case studies approaches, and that QCA thus needs to be introduced more comprehensively than other parts of a multi-method design.

Box 2.2 Multi-Method QCA: Biological Attributions of Mental Illness (Andersson and Harkness 2018)

When Do Biological Attributions of Mental Illness Reduce Stigma? Using Qualitative Comparative Analysis to Contextualize Attributions

By Matthew A. Andersson (*Baylor University*) and Sarah K. Harkness (*University of Iowa*)

In recent decades, lay and professional audiences alike have subscribed to a complex, biopsychosocial model of mental illness, which has come to supersede discrete biological, social, and psychological factors that may contribute to illness. In other words, most people usually endorse *several* intersecting beliefs surrounding the origins of mental illness — not just one or two beliefs. Beliefs about mental illness appear to be a setting where "theories do not contradict each other directly and thus do not really compete [as independent variables]"(Ragin 2008, 179).

How can we better understand how stigmatization actually works? The answer, we suspected, had to do with taking a more contingent approach to analyzing stigma. Rather than treating beliefs as isolated entities, we sought to conceptualize beliefs as constellations or "causal stories" of mental illness, which are not reducible to the elements composing them. QCA is naturally suited to uncovering highly contingent pathways between conditions and outcomes.

For us, the most challenging part about designing the QCA study, beyond learning the logic and steps involved in fs (fuzzy-set) QCA, examining best practices for QCA (e.g., Schneider and Wagemann 2010), and implementing fsQCA with *fuzzy,* a user-written Stata package (Longest and Vaisey 2008), was supplying our readers with a compelling, concise explanation of the method. How does one get quantitative scientists, who are overwhelmingly regression-minded, to intuitively appreciate what QCA reveals once "variance explained" is removed from the analytic vocabulary? We settled on the key point that covariance- or clustering-based quantitative methods treat groups of variables and their relationships as "black boxes" from which it is very difficult if not impossible to know which variables are relevant to outcomes or under which conditions they become relevant. Then, to come full-circle in our paper, we used additional, regression-based analyses, to show how QCA-based solutions explained variation in stigma even net of the singular beliefs composing them.

Our experience through the peer-review process was a bit prolonged. We went through a rejection and even a rejected revise-and-resubmit. Reviewers saw promise in our approach but reasonably had a lot of thoughtful questions about what QCA involved and how to account for our key findings. These questions thankfully moved us toward a stronger paper that could communicate with a wider intellectual audience.

When it came to anchoring our paper in a concrete social problem, the turning point for us was realizing that QCA could illuminate the "biological turn" for how mental illness is viewed. Why, if "mental illness is a disease like any other," is public acceptance not increasing? To us, there seemed a basic oversight: just because biology has become a more common explanation doesn't mean that non-biological explanations no longer are important to the public. And in fact, that is exactly what we found in our QCA study: biological and non-biological beliefs work in tandem to structure the public's desire for social distance from those perceived as mentally ill.

A Survey of Empirical Applications

To give an idea of how current QCA research *practice* looks like, as opposed to the general principles and recommendations discussed in the previous sections, this chapter closes with descriptive statistics from an original survey of 120 empirical studies. To begin with, this allows us to gauge the popularity of the different QCA variants, which are examined in more detail in Chapter 8. More importantly, the survey sheds light on the number of cases and conditions

used in applied studies from diverse academic fields in the social sciences, which we can contrast with the suggestions derived in earlier sections of this chapter.

For this survey, the *Web of Science* database was used to identify empirical QCA applications in six academic fields: business, political science, sociology, public administration, public health, and international relations. While this sample is not exhaustive in scope, the selected fields cover a broad range of social science research and academic conventions. Moreover, the selection further includes research areas where the increase in QCA studies has been most dynamic (on publication trends, see Chapter 1). Not surprisingly, publication practices vary in these fields and the prevalence of QCA in them also differs. As mentioned at the outset of this book, some fields have hundreds of QCA studies that are listed on the Web of Science, whereas for others the method is less established, and the numbers are much smaller. There are also considerable differences in the lengths of the respective articles, ranging from concise treatments of 8-10 pages to comprehensive studies of more than 30 pages in length. Naturally, this also reflects in the level of detail that is given to the methodological documentation.

To identify *current* practice in QCA applications, the search was restricted to the 20 most recent journal articles from each discipline (as of May 20, 2020), comprising a total sample of 120 empirical studies.[31] Emphasis was placed on *current* publications to assess present standards, rather than the evolution of QCA applications since the method's development in 1987. The sample was restricted to peer-reviewed journals that were listed in the Social Sciences Citation Index (SSCI). To be included in the survey, studies had to have an empirical rather than a methodological focus and use QCA as their primary documented method. Journal classification was taken as an indicator for the article's disciplinary category. Because some journals belong to several categories, classification was done in accordance with the highest-ranking category for the respective journal.

Before proceeding, one caveat is in order: given the moderate sample size, I make no claims about the representativeness of the selected studies. It should also be noted that many well-respected journals are *not* included in the SSCI, particularly non-English journals, which means there is a systematic bias in the data.[32] That said, Table 2.5 summarizes the survey results across the six disciplines. The left-hand columns show the absolute numbers of the major QCA variants (crisp-set, fuzzy-set, and multi-value QCA), whereas the right-hand columns show the median values for the number of cases, number of conditions, and the ratio of cases per condition for the observed studies.

Table 2.5 Survey of 120 QCA Studies

Academic Field		QCA Variants			Median Values		
	N	Crisp-set	Fuzzy-set	Multi-value	Cases	Conditions	Cases/Condition
Business	20	1	19	0	182	6	30.3
Sociology	20	5	15	0	30	5	4.2
International Relations	20	4	13	3	26	5	5.3
Public Health	20	5	14	1	25	6	4.7
Political Science	20	5	14	1	22	6	5.3
Public Administration	20	5	15	0	20	5	4.1
Sum/Median	120	25	90	5	26	6	5

What are the results of the survey? First, with regards to the QCA variants across the observed academic fields, it is apparent that fuzzy-set QCA has become the dominant variant. In business studies it is virtually the *only* variant that is used, but it has outpaced the other variants in all of the other fields as well. This is a surprising finding because it appears that the trend has *reversed* since the results of a comprehensive survey of 313 journal applications from 1984 to 2011, where Rihoux et al. (2013) still found that crisp-set QCA was the most popular version of QCA, with fuzzy-set QCA following in some distance. To be sure, the data might look different if we were to take a larger time span into view (for instance from 2011 to 2020), and supposedly the share of crisp-set applications would be larger then. Moreover, of the 120 studies examined, only four used multi-value QCA (Cronqvist and Berg-Schlosser 2009), which shows that this variant has remained "rather marginal", as Rihoux et al. already noted in their survey (2013, 177).[33] Not included in the table is two-step QCA (Schneider and Wagemann 2006), because it is an approach that can be combined with any QCA variant. Out of the sample of 120 studies, six articles used the two-step approach, two each in political science and sociology, and one each in business and public administration.

The second finding concerns the number of cases and conditions among the surveyed studies. Here, it is interesting to note that across fields, with the exception of business studies, the median numbers of cases and conditions used in the analysis are fairly similar – in the range of 20 to 30 cases and about 6 conditions per study. Correspondingly, the ratio of cases per condition runs from 4.1 in public administration to 5.3 in political science and international relations. At the far end are business studies with a median of 30.3 cases per condition. With these median values, the surveyed articles are *above* the recommended threshold of cases per conditions, as summarized in Table 2.2 earlier in this chapter. However, it should be clear that these numbers refer to the *median* across the sampled studies, so individual studies diverge from these values. Finally, there is an evident break between the business field and the other academic disciplines, with the former being characterized by large-*N* fuzzy-set QCA with considerably more cases per study than commonly used in other areas of the social sciences.

Notes

[1] Toshkov (2016, 1).

[2] For texts on general aspects of research design in the social sciences, see Blatter and Haverland (2012); Brady and Collier (2010); George and Bennett (2005); Gerring (2012); Goertz and Mahoney (2012); Gschwend and Schimmelfennig (2007); King et al. (1994); Ragin and Becker (1992); Rohlfing (2012); Toshkov (2016). On QCA research design, see Thomann and Maggetti (2020).

[3] Chapter 8 discusses some proposals made towards temporal QCA, as a formal way of including sequence in the analytical procedure.

[4] For further contributions to this debate, see, among many others, Tarrow (1995), George and Bennett (2005), Rihoux and Grimm (2006), Ragin (2008), Collier et al. (2010), Mahoney (2010), Goertz and Mahoney (2012), and Cooper et al. (2012). Some of the discussions have been continued in various issues of *Qualitative and Multi-Method Research* of the American Political Science Association.

[5] Apart from King et al. (1994, Ch. 1), Schmitter (2008), and Gerring (2012, Ch. 2), see also the helpful discussion of types of research and research questions in Toshkov (2016, Ch. 2).

[6] This draws on the distinction between x-centered and y-centered research that was established by Ganghof (2005).

[7] In his book on multi-method research, Goertz (2017, Ch. 2) discusses published examples of how QCA is used to explore causal mechanisms.

[8] Cast at a more general level than other accounts, my taxonomy differs from those who include five or even six types of different uses of QCA (Berg-Schlosser et al. 2009, 15; Ragin and Rihoux 2004, 6; Schneider and Wagemann 2010, 400). The distinguishing criterion for these is whether a study's primary aim is descriptive or explanatory. Apart from that, some suggested categories such as "summarizing data" and "checking coherence of data" appear virtually indistinguishable in applied settings.

[9] Apart from a handful exceptions, nearly all of the 120 QCA applications surveyed for this chapter used the method to test theoretical expectations in one way or another. The results of the survey are presented in this chapter's final section.

[10] On the history of statistical hypothesis testing, see Haig (2018, Ch. 3).

[11] Formal differences between these perspectives are examined in Thiem et al. (2016).

[12] Other examples for the use of INUS hypotheses include, Ide (2015), Mello (2020), Oppermann and Brummer (2020), and Wurster and Hagemann (2018). On the concept, see Chapter 2.

[13] The terms population and universe of cases are used interchangeably (Seawright and Collier 2010, 357).

[14] Case selection and case study methods have spawned an extensive literature, see Blatter and Haverland (2012); Eckstein (1975); George and Bennett (2005); Gerring (2004; 2007); Goertz and Mahoney (2006); Levy (2008); Ragin and Becker (1992); Rohlfing (2012); Seawright and Gerring (2008).

[15] For illustrated discussions of scope conditions, see Goertz and Mahoney (2006) and Goertz (2017). More generally on constituting populations of cases, see Ragin (2000, Ch. 2 & 7); (2006).

[16] A discussion of sources of bias in case selection is provided in Geddes (2007, Ch. 3). See also Van Evera (1997, 46).

[17] For more general considerations of sampling, see Frankfort-Nachmias and Nachmias (2008, Ch. 8) and Toshkov (2016, 130-34).

[18] A clear exception are studies in business, management, and economics, which tend to have considerably higher numbers of cases (see the discussion of the survey results at the end of this section).

[19] Consistency and other measures of fit are discussed in Chapter 6. Here, it suffices to grasp the general logic of the relationship between cases and conditions.

[20] There is no general answer to these questions. For some studies, it may be feasible to adapt the research design to increase the empirical scope, whereas others might be constrained by prior decisions.

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

[22] The sole exception is two-step QCA, which rests on a theoretical distinction between "remote" and "proximate" conditions (Schneider and Wagemann 2006). Examples are discussed in Chapter 8.

[23] Notably, this is not a technical limitation, because the QCA package for R can include up to 18 conditions (Duşa 2019, 202). However, for applied research, a study with 262,144 truth table rows would not be feasible or allow for a shallow analysis at best.

[24] Goertz (2017, 46-49) discusses the integrated model provided in Mello (2012).

[25] See also the separate info box by Maria Brockhaus and colleagues, who share insights on the larger research project behind their QCA studies (Brockhaus et al. 2017).

[26] While these terms are at times used synonymously, they reflect different understandings of how methods and data from different sources ought to be combined. For this book, the term *multi-method research* is used, to designate the combination of QCA with case studies.

[27] Along such lines, George and Bennett (2005, 163) argue that it is necessary to combine QCA with process tracing.

[28] Others have suggested that combining methods holds no inherent advantages over the use of a single method (Ahmed and Sil 2009; Coppedge 2009; Kuehn and Rohlfing 2009).

[29] Though less common, QCA has been combined with large-*N* statistical analyses (e.g. Ahn and Lee 2012; Ide 2018; Karlas 2012). In this book, see also the info box by Tobias Ide, reflecting on his study on environmental peacekeeping (Ide 2018).

[30] Schneider and Rohlfing (2013, 585) further denote cases that show the outcome without membership in the solution as "deviant cases for coverage". These are cases that call for a different explanation.

[31] Thanks to Teslin Augustine for research assistance with the survey.

[32] On the coverage of the Web of Science and alternative databases, see Gerring et al. (2020).

[33] For an example of mvQCA, see the info box by Pablo Castillo-Ortiz on his study on judicial councils (Castillo Ortiz 2017). See also the discussion in Chapter 8.

References

Ahmed, Amel, and Rudra Sil. 2009. "Is Multi-Method Research Really "Better"?" *Qualitative & Multi-Method Research* 7 (2): 2-6.

Ahn, S., and S.S. Lee. 2012. "Explaining Korean Welfare State Development with New Empirical Data and Methods." *Asian Social Work and Policy Review* 6 (2): 67–85.

Amenta, Edwin, and J. D. Poulsen. 1994. "Where to Begin: A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis." *Sociological Methods & Research* 23 (1): 22–53.

Andersson, Matthew A., and Sarah K. Harkness. 2018. "When Do Biological Attributions of Mental Illness Reduce Stigma? Using Qualitative Comparative Analysis to Contextualize Attributions." *Society and Mental Health* 8 (3): 175-94.

Anjum, Rani Lill, and Stephen Mumford. 2018. *Causation in Science*. New York: Oxford University Press.

Avdagic, Sabina. 2010. "When Are Concerted Reforms Feasible? Explaining the Emergence of Social Pacts in Western Europe." *Comparative Political Studies* 43 (5): 628–57.

Beach, Derek. 2018. "Achieving Metholodogical Alignment When Combining QCA and Process Tracing in Practice." *Sociological Methods & Research* 47 (1): 64-99.

———. 2020. "Multi-Method Research in the Social Sciences: A Review of Recent Frameworks and a Way Forward." *Government and Opposition* 55 (1): 163-82.

Beach, Derek, and Rasmus Brun Pedersen. 2013. *Process-Tracing Methods: Foundations and Guidelines*. 1st ed. Ann Arbor, MI: University of Michigan Press.

———. 2019. *Process-Tracing Methods: Foundations and Guidelines*. 2nd ed. Ann Arbor, MI: University of Michigan Press.

Beach, Derek, Rasmus Brun Pedersen, and Markus B. Siewert. 2019. "Case Selection and Nesting of Process-Tracing Case Studies." In *Process-Tracing Methods: Foundations and Guidelines*, edited by Beach, Derek and Rasmus Brun Pedersen. Ann Arbor, MI: University of Michigan Press, 89-154.

Beach, Derek, and Ingo Rohlfing. 2018. "Integrating Cross-case Analyses and Process Tracing in Set-Theoretic Research: Strategies and Parameters of Debate." *Sociological Methods & Research* 47 (1): 3-36.

Berg-Schlosser, Dirk, and Gisèle De Meur. 2009. "Comparative Research Design: Case and Variable Selection." In *Configurational Comparative Methods*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 19-32.

Berg-Schlosser, Dirk, Gisèle De Meur, Benoît Rihoux, and Charles C. Ragin. 2009. "Qualitative Comparative Analysis (QCA) as an Approach." In *Configurational Comparative Methods*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 1–18.

Binder, Martin. 2015. "Paths to Intervention: What Explains the UN's Selective Response to Humanitarian Crises?" *Journal of Peace Research* 52 (6): 712-26.

Blatter, Joachim, and Markus Haverland. 2012. *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.

Boogaerts, Andreas. 2018. "Beyond Norms: A Configurational Analysis of the EU's Arab Spring Sanctions." *Foreign Policy Analysis* 14 (3): 408-28.

Brady, Henry E., and David Collier, ed. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham: Rowman & Littlefield.

———, ed. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham: Rowman & Littlefield.

Brockhaus, Maria, Kaisa Korhonen-Kurki, Jenniver Sehring, Monica Di Gregorio, Samuel Assembe-Mvondo, Hermann Kambire, Felicien Kengoum, Demetrius Kweka, Mary Menton, Moira Moeliono, Naya Sharma Paudel, Thuy Thu Pham, Ida Aju Pradnja Resosudarmo, Almeida Sitoe, Sven Wunder, and Mathurin Zida. 2017. "REDD+, Transformational Change and the Promise of Performance-Based Payments: A Qualitative Comparative Analysis." *Climate Policy* 17 (6): 708-30.

Caporaso, James A. 1995. "Research Design, Falsification, and the Qualitative-Quantitative Divide." *American Political Science Review* 89 (2): 457-60.

Castillo Ortiz, Pablo José. 2017. "Councils of the Judiciary and Judges' Perceptions of Respect to Their Independence in Europe." *Hague Journal on the Rule of Law* 9 (2): 315-36.

Collier, David, Henry E. Brady, and Jason Seawright. 2010. "Critiques, Responses, and Trade-Offs: Drawing Together the Debate." In *Rethinking Social Inquiry*, edited by Brady, Henry E. and David Collier. Lanham, MD: Rowman & Littlefield, 125–59.

Collier, David, Jody Laporte, and Jason Seawright. 2008. "Typologies: Forming Concepts and Creating Categorical Variables." In *The Oxford Handbook of Political Methodology*, edited by Box-Steffensmeier, Janet M., Henry E. Brady and David Collier. Oxford: Oxford University Press, 152–73.

Cooper, Barry, Judith Glaesser, Roger Gomm, and Martyn Hammersley. 2012. *Challenging the Qualitative-Quantitative Divide: Explorations in Case-Focused Causal Analysis*. London: Continuum.

Coppedge, Michael. 2009. "Speedbumps on the Road to Multi-Method Consensus in Comparative Politics." *Qualitative & Multi-Method Research* 7 (2): 15-17.

Cronqvist, Lasse, and Dirk Berg-Schlosser. 2009. "Multi-Value QCA (mvQCA)." In *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, edited by Rihoux, Benoît and Charles C. Ragin. Los Angeles: Sage, 69-86.

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

Eckstein, Harry. 1975. "Case Study and Theory in Political Science." In *Handbook of Political Science*, edited by Greenstein, Fred I. and Nelson W. Polsby. Reading: Addison-Wesley, 79–137.

Elman, Colin. 2005. "Explanatory Typologies in Qualitative Studies of International Politics." *International Organization* 59: 293–326.

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. Los Angeles: Sage.

Frankfort-Nachmias, Chava, and David Nachmias. 2008. *Research Methods in the Social Sciences*. 7th ed. New York, NY: Worth Publishers.

Ganghof, Steffen. 2005. "Kausale Perspektiven in der vergleichenden Politikwissenschaft: X-zentrierte und Y-zentrierte Forschungsdesigns." In *Vergleichen in der Politikwissenschaft*, edited by Kropp, Sabine and Michael Minkenberg. Wiesbaden: VS Verlag, 76-93.

Geddes, Barbara. 2007. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor, MI: University of Michigan Press.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge: MIT Press.

Gerring, John. 2004. "What Is a Case Study and What Is It Good for?" *American Political Science Review* 98 (2): 341-54.

———. 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.

———. 2012. *Social Science Methodology: A Unified Framework*. 2nd. Cambridge: Cambridge University Press.

Gerring, John, Sebastian Karcher, and Brendan Apfeld. 2020. "Impact Metrics." In *The Production of Knowledge: Enhancing Progress in Social Science*, edited by Elman, Colin, John Gerring and James Mahoney. Cambridge: Cambridge University Press, 371-99.

Goertz, Gary. 2003a. "Cause, Correlation, and Necessary Conditions." In *Necessary Conditions*, edited by Goertz, Gary and Harvey Starr. Lanham: Rowman & Littlefield, 47–64.

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

———. 2017. *Multimethod Research, Causal Mechanisms, and Case Studies*. Princeton, NJ: Princeton University Press.

Goertz, Gary, and James Mahoney. 2006. "Negative Case Selection: The Possibility Principle." In *Social Science Concepts*, edited by Goertz, Gary. Princeton, NJ: Princeton University Press, 177–210.

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

Gschwend, Thomas, and Frank Schimmelfennig, ed. 2007. *Research Design in Political Science.* Basingstoke: Palgrave Macmillan.

Haesebrouck, Tim. 2017. "NATO Burden Sharing in Libya: A Fuzzy Set Qualitative Comparative Analysis." *Journal of Conflict Resolution* 61 (10): 2235-61.

Haig, Brian D. 2018. *The Philosophy of Quantitative Methods*. New York: Oxford University Press.

Held, Tobias, and Lasse Gerrits. 2019. "One the Road to Electrification - A Qualitative Comparative Analysis of Urban E-Mobility Policies in 15 European Cities." *Transport Policy* 81: 12-23.

Hörisch, Felix. 2013. "Fiscal Policy in Hard Times: A Fuzzy-Set QCA of Fiscal Policy Reactions to the Financial Crisis." *Zeitschrift für Vergleichende Politikwissenschaft* 7: 117-41.

Ide, Tobias. 2015. "Why Do Conflicts over Scarce Renewable Resources Turn Violent? A Qualitative Comparative Analysis." *Global Environmental Change* 33: 61-70.

———. 2018. "Does Environmental Peacemaking between States Work? Insights on Cooperative Environmental Agreements and Reconciliation in International Rivalries." *Journal of Peace Research* 55 (3): 351-65.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory Meets Scientific Practice*. Oxford: Oxford University Press.

Kahwati, Leila C., and Heather L. Kane. 2020. *Qualitative Comparative Analysis in Mixed Methods Research and Evaluation*. Los Angeles Sage.

Karlas, Jan. 2012. "National Parliamentary Control of EU Affairs: Institutional Design after Enlargement." *West European Politics* 35 (5): 1095–113.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Kirchherr, Julian, Mats-Philip Ahrenshop, and Katrina Charles. 2019. "Resettlement Lies: Suggestive Evidence from 29 Large Dam Projects." *World Development* 114: 208-19.

Kuehn, David, Aurel Croissant, Jil Kamerling, Hans Lueders, and André Strecker. 2017. "Conditions of Civilian Control in New Democracies: An Empirical Analysis of 28 'Third Wave' Democracies." *European Political Science Review* 9 (3): 425-47.

Kuehn, David, and Ingo Rohlfing. 2009. "Does It, Really? Measurement Error and Omitted Variables in Multi-Method Research." *Qualitative & Multi-Method Research* 7 (2): 18-22.

Kvist, Jon. 1999. "Welfare Reform in the Nordic Countries in the 1990s: Using Fuzzy-Set Theory to Assess Conformity to Ideal Types." *Journal of European Social Policy* 9 (3): 231-52.

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

Levy, Jack S. 2008. "Case Studies: Types, Designs, and Logics of Inference." *Conflict Management and Peace Science* 25 (1): 1–18.

Lieberman, Evan S. 2005. "Nested Analysis as a Mixed-Method Strategy for Comparative Research." *American Political Science Review* 99 (3): 435–52.

———. 2015. "Nested Analysis: Toward the Integration of Comparative-Historical Analysis with other Social Science Methods." In *Advances in Comparative-Historical Analysis*, edited by Mahoney, James and Kathleen Thelen. Cambridge: Cambridge University Press, 240-63.

Lieberson, Stanley. 1991. "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70 (2): 307–20.

Lilliefeldt, Emelie. 2012. "Party and Gender in Western Europe Revisited: A Fuzzy-Set Qualitative Comparative Analysis of Gender-Balanced Parliamentary Parties." *Party Politics* 18 (2): 193–214.

Longest, Kyle C., and Stephen Vaisey. 2008. "Fuzzy: A Program for Performing Qualitative Comparative Analyses (QCA) in Stata." *The Stata Journal* 8 (1): 79-104.

Maat, Eelco van der. 2011. "Sleeping Hegemons: Third-Party Intervention Following Territorial Integrity Transgressions." *Journal of Peace Research* 48 (2): 201–15.

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

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2019. "Polity IV Project: Dataset User's Manual." Vienna, VA: Center for Systemic Peace.

Marx, Axel. 2006. "Towards More Robust Model Specifiation in QCA: Results from a Methodological Experiment." *COMPASSS Working Papers*: 1-25.

Marx, Axel, and Adrian Duşa. 2011. "Crisp-Set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification." *Methodological Innovations Online* 6 (2): 102-48.

Meegdenburg, Hilde van, and Patrick A. Mello. forthcoming. *Case-Based Methods: Combining Process Tracing and Qualitative Comparative Analysis*. London: Palgrave Macmillan.

Mello, Patrick A. 2012. "Parliamentary Peace or Partisan Politics? Democracies' Participation in the Iraq War." *Journal of International Relations and Development* 15 (3): 420–53.

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

Merton, Robert K. 1958. *Social Theory and Social Structure*. New York: Free Press.

Meuer, Johannes, and Christian Rupietta. 2017. "A Review of Integrated QCA and Statistical Analyses." *Quality & Quantity* 51 (5): 2063-83.

Nelson, Jennifer M. 2017. "Pathways to Green(er) Pastures: Reward Bundles, Human Capital, and Turnover Decisions in a Semi-Profession." *Qualitative Sociology* 40 (1): 23-57.

Oppermann, Kai, and Klaus Brummer. 2020. "Who Gets What in Foreign Affairs? Explaining the Allocation of Foreign Ministries in Coalition Governments." *Government and Opposition* 55 (2): 241-59.

Pattyn, Valérie, Priscilla Álamos-Concha, Bart Cambré, Benoît Rihoux, and Benjamin Schalembier. 2020. "Policy Effectiveness through Configurational and Mechanistic Lenses: Lessons for Concept Development." *Journal of Comparative Policy Analysis: Research and Practice* online first (doi:10.1080/13876988.2020.1773263): 1-18.

Pinfari, Marco. 2011. "Time to Agree: Is Time Pressure Good for Peace Negotiations?" *Journal of Conflict Resolution* 55 (5): 683–709.

Pullum, Amanda. 2016. "Social Movements, Strategic Choice, and Recourse to the Polls." *Mobilization: An International Journal* 21 (2): 177-92.

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

———. 2006. "How to Lure Analytic Social Science Out of the Doldrums: Some Lessons from Comparative Research." *International Sociology* 21 (5): 633-46.

———. 2008. *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 Howard S. Becker, ed. 1992. *What Is A Case? Exploring the Foundations of Social Inquiry*. New York: Cambridge University 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., and Benoît Rihoux. 2004. "Qualitative Comparative Analysis (QCA): State of the Art and Prospects." *Qualitative Methods* 2 (2): 3-13.

Rihoux, Benoît, Priscilla Álamos-Concha, Damien Bol, Axel Marx, and Ilona Rezsöhazy. 2013. "From Niche to Mainstream Method? A Comprehensive Mapping of QCA Applications in Journal Articles from 1984 to 2011." *Political Research Quarterly* 66 (1): 175-84.

Rihoux, Benoît, Priscilla Álamos-Concha, and Bojana Lobe. 2021. "Qualitative Comparative Analysis (QCA): An Integrative Approach Suited for Diverse Mixed Method and Multimethod Research Strategies." In *The Routledge Reviewer's Guide to Mixed Methods Analysis*, edited by Onwuegbuzie, Anthony J. and R. Burke Johnson. London: Routledge.

Rihoux, Benoît, and Heike Grimm, ed. 2006. *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide.* Boston, MA: Springer.

Rihoux, Benoît, and Bojana Lobe. 2009. "The Case for Qualitative Comparative Analysis (QCA): Adding Leverage for Thick Cross-Case Comparison." In *The Sage Handbook of Case-Based Methods*, edited by Byrne, David and Charles C. Ragin. Los Angeles: Sage, 222–42.

Rihoux, Benoît, Charles C. Ragin, Sakura Yamasaki, and Damien Bol. 2009. "Conclusions - The Way(s) Ahead." In *Configurational Comparative Methods*, edited by Rihoux, Benoît and Charles C. Ragin. Thousand Oaks: Sage, 167–77.

Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. Basingstoke: Palgrave Macmillan.

Rohlfing, Ingo, and Carsten Q. Schneider. 2018. "A Unifying Framework for Causal Analysis in Set-Theoretic Multimethod Research." *Sociological Methods & Research* 47 (1): 37-63.

Schmitter, Philippe C. 2008. "The Design of Social and Political Research." In *Approaches and Methodologies in the Social Sciences*, edited by Della Porta, Donatella and Michael Keating. Cambridge: Cambridge University Press, 263–95.

Schneider, Carsten Q., and Ingo Rohlfing. 2013. "Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research." *Sociological Methods & Research* 42 (4): 559-97.

———. 2019. "Set-theoretic Multimethod Research: The Role of Test Corridors and Conjunctions for Case Selection." *Swiss Political Science Review* 25 (3): 253-75.

Schneider, Carsten Q., Barbara Vis, and Kendra Koivu. 2019. "Set-Analytic Approaches, Especially Qualitative Comparative Analysis (QCA)." APSA Section for Qualitative and Multi-Method Research, Final Report of the Qualitative Transparency Deliberations Working Group III.4.

Schneider, Carsten Q., and Claudius Wagemann. 2006. "Reducing Complexity in Qualitative Comparative Analysis (QCA): Remote and Proximate Factors and the Consolidation of Democracy." *European Journal of Political Research* 45 (5): 751–86.

———. 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.

Seawright, Jason. 2016. *Multi-Method Social Science: Combining Qualitative and Quantitative Tools*. Cambridge: Cambridge University Press.

Seawright, Jason, and David Collier. 2010. "Glossary." In *Rethinking Social Inquiry*, edited by Brady, Henry E. and David Collier. Lanham: Rowman & Littlefield, 313-59.

Seawright, Jason, and John Gerring. 2008. "Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options." *Political Research Quarterly* 61 (2): 294–308.

Sessions, Roger. 1950. "How A 'Difficult' Composer Gets That Way." *New York Times* 8 January.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.

Tarrow, Sidney. 1995. "Bridging the Quantitative-Qualitative Divide in Political Science." *American Political Science Review* 89 (2): 471-74.

Thiem, Alrik, Michael Baumgartner, and Damien Bol. 2016. "Still Lost in Translation! A Correction of Three Misunderstandings Between Configurational Comparativists and Regressional Analysts." *Comparative Political Studies* 49 (6): 742-74.

Thomann, Eva, and Martino Maggetti. 2020. "Designing Research With Qualitative Comparative Analysis (QCA): Approaches, Challenges, and Tools." *Sociological Methods & Research* 49 (2): 356-86.

Toshkov, Dimiter. 2016. *Research Design in Political Science*. Basingstoke: Palgrave Macmillan.

Van Evera, Stephen. 1997. *Guide to Methods for Students of Political Science*. 1. . New York: Cornell University Press.

Vis, Barbara. 2010. *Politics of Risk-Taking: Welfare State Reform in Advanced Democracies*. Amsterdam: Amsterdam University Press.

———. 2011. "Under Which Conditions Does Spending on Active Labor Market Policies Increase? An fsQCA Analysis of 53 Governments between 1985 and 2003." *European Political Science Review* 3 (2): 229–52.

Wurster, Stefan, and Christian Hagemann. 2018. "Two Ways to Success: Expansion of Renewable Energies in Comparison Between Germany's Federal States." *Energy Policy* 119: 610-19.