The bare-bone basics of crisp-set QCA

Step 1: Identify relevant cases and causal conditions

1-1. Identify the **outcome** that you are interested in and the cases that exemplify this outcome. Learn as much as you can about these “**positive**” cases.

1-2. Based on #1, identify **negative** cases—those that might seem to be candidates for the outcome but nevertheless failed to display it (“negative” cases). Together #1 and #2 constitute the **set of cases relevant to the analysis**.

1-3. Again based on #1, and relevant theoretical and substantive knowledge, identify the major **causal conditions** relevant to the outcome. Often, it is useful to think in terms of different causal “recipes”—the various combinations of conditions that might generate the outcome.

1-4. Try to **streamline** the causal conditions as much as possible. For example, combine two conditions into one when they seem “substitutable.”

Example:

1. Identify positive instances of mass protest against austerity measures mandated by the International Monetary Fund (IMF) as conditions for debt renegotiation (“conditionality”). Peru, Argentina, Tunisia, . . .

2. Identify negative cases: for example, debtor countries that were also subject to strict IMF conditionality, but nevertheless did not experience mass protest. Mexico, Costa Rica, . . .

3. Identify relevant causal conditions: severity of austerity measures, degree of debt, living conditions, consumer prices, prior levels of political mobilization, government corruption, union strength, trade dependence, investment dependence, urbanization and other structural conditions relevant to protest mobilization. One recipe might be severe austerity measures combined with government corruption, rapid consumer price increases and high levels of prior political mobilization. Again, knowledge of positive cases is very important.

4. Streamlining example: Based on case knowledge, the researcher might surmise that high trade dependence and high investment dependence are substitutable manifestations of international economic dependence and therefore create a single condition from these two, using logical “or.”

Step 2: Construct the data set

QCA data sets have the standard format: Cases define the rows, and aspects of cases define the columns.

However, it is important to remember that the columns are sets and that the relevant sets must be carefully conceptualized and labeled.

The cells of the data spreadsheet are filled with set membership scores. When using crisp sets, there are only two possible values: 1 (member) and 0 (nonmember).

It is important to specify set membership criteria as carefully as possible and make sure that the membership scores that are assigned are faithful to the set label and how it is conceptualized.

Very often it is useful to use to same evidence (e.g., parental income) as a basis for more than one set. For example, the set of individual with high-income parents has different membership than the set of individual with not-low income parents. The former is a subset of the latter. These multiple conceptualizations based on the same source variable (e.g., parental income) can be used in the same analysis.

Here's a sample data set with crisp sets (N = 68):

CaseID	PriorMob	SevereAus	GovCor	PriceRis	Protest
a	0	0	1	0	0
b	0	0	1	0	0
c	0	0	1	0	0
d	0	0	1	0	0
e	0	0	1	1	0
f	0	0	1	1	0
g	0	0	1	1	0
h	0	0	1	1	0
i	0	0	1	1	0
j	0	0	1	1	1
k	0	1	0	1	1
l	0	1	0	1	1
m	0	1	0	1	1
n	0	1	0	1	1
o	0	1	1	1	1
p	0	1	1	1	1
q	0	1	1	1	1
r	0	1	1	1	1
s	0	1	1	1	1
t	1	0	0	0	0
u	1	0	0	0	0
v	1	0	0	0	0
w	1	0	0	1	0
x	1	0	0	1	0
y	1	0	0	1	0
z	1	0	0	1	0
etc.	1	0	0	1	0
	1	0	0	1	0
	1	0	0	1	0
	1	0	0	1	1
	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0

	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0
	1	0	1	0	0
	1	1	0	0	1
	1	1	0	0	0
	1	1	0	0	0
	1	1	0	0	0
	1	1	0	0	0
	1	1	0	0	0
	1	1	0	1	1
	1	1	0	1	1
	1	1	0	1	1
	1	1	0	1	1
	1	1	0	1	1
	1	1	0	1	1
	1	1	1	0	0
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	0	1
	1	1	1	1	0
	1	1	1	1	0
	1	1	1	1	0
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1

	1	1	1	1	1
	1	1	1	1	1

Step 3: Test for necessary conditions

The core of QCA is truth table analysis, which looks for causal combinations ("recipes") that are sufficient for the outcome. Before conducting a truth table analysis, however, it is useful to see if any of the causal conditions might be considered **necessary (but not sufficient)** conditions for the outcome.

If a causal condition is shared as an antecedent condition by instances of the outcome, then instances of the outcome constitute a **subset** of instances of the causal condition. (Example: having an advanced degree is a shared antecedent condition for many professional occupations; still, there are plenty of Ph.D.s driving taxis, so it's not a sufficient condition).

With crisp sets, the simplest way to assess necessary conditions is the use QCA's crosstabs procedure and ask for row percentages. Look across the row showing the presence of the outcome; if all instances (or almost all instances) agree in displaying a particular causal condition, then that condition might be interpreted as a necessary condition (shared antecedent condition).

You can also look across the rows of a crosstabulation to see if a causal condition is a shared antecedent condition. If it is, then the (0,1) cell (i.e., the first cell) will be empty and the (1,1) cell (i.e., the second) will be well populated.

Using the data set presented above we find:

```
N                |
Row %            |
                 |priormob
protest          |        0 |        1 |
-----------------+----------+----------+
1                |       10 |       22 |         32
                 |     31.3 |     68.8 |
-----------------+----------+----------+
0                |        9 |       27 |         36
                 |     25.0 |     75.0 |
-----------------+----------+----------+
                         19         49
Total N                  68
Missing                   0

N                |
Row %            |
                 |severeaus
protest          |        0 |        1 |
-----------------+----------+----------+
1                |        2 |       30 |         32
                 |      6.3 |     93.8 |
-----------------+----------+----------+
0                |       29 |        7 |         36
                 |     80.6 |     19.4 |
-----------------+----------+----------+
                         31         37
Total N                  68
Missing                   0
```

N Row % protest	govcorupt 0	 1	
1	12 37.5	20 62.5	32
0	15 41.7	21 58.3	36
	27	41	
Total N	68		
Missing	0		

N Row % protest	pricerise 0	 1	
1	7 21.9	25 78.1	32
0	24 66.7	12 33.3	36
	31	37	

The key calculation is the percentage of the cases with the outcome (first row in these tables) that also display the causal condition (cell 2, in these tables). The row percentage should be 100% or very close to 100%.

The causal condition “severeaus” (severe austerity) is close enough (at 93.9%) to be considered a candidate for necessity. It is a “widely shared” antecedent condition. When evaluating possible necessary conditions, I look for consistency scores of at least 90%.

It is important to assess whether the condition makes sense as a necessary condition.

Remember that a necessary condition is a superset of the outcome, while a sufficient condition (or combination of conditions) is a subset.

In general, it is useful to know which conditions are supersets of the outcome because they may have special practical value of theoretic status.

Step 4: Construct the truth table and resolve contradictions

4-1. Construct a “truth table” based on the causal conditions specified in step 1 or some reasonable subset of these conditions (e.g., using a recipe that seems especially promising). A truth table sorts cases by the combinations of causal conditions they exhibit. All logically possible combinations of conditions are considered, even those without empirical instances.

4-2. Assess the consistency of the cases in each row with respect to the outcome: Do they uniformly share the outcome? A simple measure of consistency for crisp sets is the percentage of cases in each row of the truth table displaying the outcome. Consistency scores of either 1 or 0 indicate perfect consistency for a given row. A score of 0.50 indicates perfect inconsistency.

4-3. Identify contradictory rows. Technically, a contradictory row is any row with a consistency score that is not equal to 1 or 0. However, it is sometimes reasonable to relax this standard, for example, if an inconsistent case in a given row can be explained by its specific circumstances.

4-4. Compare cases within each contradictory rows. If possible, identify decisive differences between positive and negative cases, and then revise the truth table accordingly.

Example, using one possible recipe as a starting point:

Row#	Prior mobiliz.?	Severe austerity?	Gov't corrupt?	Rapid price rise?	Cases w/ protest?	Cases w/o protest	Consis-tency
1	0 (no)	0 (no)	0 (no)	0 (no)	0	0	??
2	0 (no)	0 (no)	0 (no)	1 (yes)	0	0	??
3	0 (no)	0 (no)	1 (yes)	0 (no)	0	4	0.0
4	0 (no)	0 (no)	1 (yes)	1 (yes)	1	5	0.167
5	0 (no)	1 (yes)	0 (no)	0 (no)	0	0	??
6	0 (no)	1 (yes)	0 (no)	1 (yes)	4	0	1.0
7	0 (no)	1 (yes)	1 (yes)	0 (no)	0	0	??
8	0 (no)	1 (yes)	1 (yes)	1 (yes)	5	0	1.0
9	1 (yes)	0 (no)	0 (no)	0 (no)	0	3	0.0
10	1 (yes)	0 (no)	0 (no)	1 (yes)	1	7	0.125
11	1 (yes)	0 (no)	1 (yes)	0 (no)	0	10	0.0
12	1 (yes)	0 (no)	1 (yes)	1 (yes)	0	0	??
13	1 (yes)	1 (yes)	0 (no)	0 (no)	1	5	0.167
14	1 (yes)	1 (yes)	0 (no)	1 (yes)	6	0	1.0
15	1 (yes)	1 (yes)	1 (yes)	0 (no)	6	2	0.75
16	1 (yes)	1 (yes)	1 (yes)	1 (yes)	8	0	1.0

Table Notes:

a. This table has five rows *without* cases (1, 2, 5, 7, 12). In QCA, these rows are labeled “remainder rows.” Having remainders is known as “limited diversity.”

b. There are seven noncontradictory rows, three that are uniform in not displaying the outcome (consistency = 0.0; rows 3, 9, 11) and four that are uniform in displaying the outcome (consistency = 1.0; rows 6, 8, 14, 16).

c. The remaining four rows are contradictory. Three are close to 0.0 (rows 4, 10, 13), and one is close to 1.0 (row 15).

d. Suppose that the three (unexpected) positive cases (one each in rows 4, 10, 13) are all cases of contagion—a neighboring country with severe IMF protest spawned protest in these three countries. These contradictory cases can be explained using case knowledge, showing that these instances are irrelevant to the recipe in question. Thus, these three cases can be safely set aside.

e. Suppose the comparison of the positive and negative cases in row 15, reveals that the (unexpected) negative cases all had severely repressive regimes. This pattern suggests that having a not-severely-repressive regime is part of the recipe and that the recipe has five key conditions, not four. The revised truth table follows. (To simplify the presentation, only rows with cases are shown.)

Prior mobiliz.?	Severe austerity?	Gov’t corrupt?	Rapid price rise?	Not re-pressive?	Cases w/ protest?	Cases w/o protest	Consis-tency
0 (no)	0 (no)	1 (yes)	0 (no)	0 (no)	0	4	0.0
0 (no)	0 (no)	1 (yes)	1 (yes)	0 (no)	0	5	0.0
0 (no)	1 (yes)	0 (no)	1 (yes)	0 (no)	4	0	1.0
0 (no)	1 (yes)	1 (yes)	1 (yes)	1 (yes)	5	0	1.0
1 (yes)	0 (no)	0 (no)	0 (no)	0 (no)	0	3	0.0
1 (yes)	0 (no)	0 (no)	1 (yes)	1 (yes)	0	7	0.0
1 (yes)	0 (no)	1 (yes)	0 (no)	0 (no)	0	10	0.0
1 (yes)	1 (yes)	0 (no)	0 (no)	1 (yes)	0	5	0.0
1 (yes)	1 (yes)	0 (no)	1 (yes)	0 (no)	6	0	1.0
1 (yes)	1 (yes)	1 (yes)	0 (no)	1 (yes)	6	0	1.0
1 (yes)	1 (yes)	1 (yes)	0 (no)	0 (no)	0	2	0.0
1 (yes)	1 (yes)	1 (yes)	1 (yes)	1 (yes)	8	0	1.0

Notice that there are no contradictions. The three cases of contagion have been removed, and the contradictory cases in row 15 of the previous table have been resolved.

The “full” version of this truth table would have 32 rows and 20 “remainder rows” (causal combinations lacking cases).

Step 5: Analyze the truth table

fsQCA software can be used to analyze truth tables like the two just shown. The goal of the analysis is to specify the different combinations of conditions linked to the selected outcome, based on the features of the positive cases that consistently distinguish them from the negative cases.

5-1. The first part of the fsQCA algorithm compares rows of the truth table to identify matched pairs. For example, these two rows both have the outcome, but differ by ONLY ONE causal condition, providing an experiment-like contrast:

Prior mobiliz.?	Severe austerity?	Gov't corrupt?	Rapid price rise?	Not re-pressive?	Cases w/ protest?	Cases w/o protest	Consis-tency
1 (yes)	1 (yes)	1 (yes)	0 (no)	1 (yes)	6	0	1.0
1 (yes)	1 (yes)	1 (yes)	1 (yes)	1 (yes)	8	0	1.0

This paired comparison indicates that if prior mobilization, severe IMF austerity, government corruption, and a nonrepressive regime coincide, it doesn't matter whether there are also rapid price increases; protest will still erupt. The term that differs is eliminated, and a single, simpler row replaces the two rows shown.

5-2. This process of bottom-up paired comparison continues until no further simplification is possible. Only rows with the outcome are paired; only one condition may differ in each paired comparison. The one that differs is eliminated.

5-3. Rows without cases (“remainders”) also may be used to aid the process of simplifying the patterns. For example, there are no instances of the second row listed below. However, based on substantive and theoretical knowledge, it is reasonable to speculate that if such cases existed, they would be positive instances of IMF protest, just like the empirical cases they are paired with:

Prior mobiliz.?	Severe austerity?	Gov’t corrupt?	Rapid price rise?	Not Re-pressive?	Cases w/ protest?	Cases w/o protest	Consis-tency
0 (no)	1 (yes)	0 (no)	1 (yes)	0 (no)	4	0	1.0
0 (yes)	1 (yes)	1 (yes)	1 (yes)	0 (no)	0	0	??

The reasoning is as follows: (1) the remainder case resembles the empirical cases above it in every respect except one; (2) the one difference (the remainder case has government corruption) involves a condition that should only make IMF protest *more likely*; (3) therefore, the remainder case, if it existed, would display IMF protest, just like the empirical cases. This pairing allows the production of a logically simpler configuration, eliminating the absence of government corruption as a possible (INUS) ingredient.

This example shows how fsQCA incorporates a form of counterfactual analysis that parallels practices in qualitative case-oriented research.

5-4. The process of paired comparisons culminates in a list of causal combinations linked to the outcome. fsQCA then selects the smallest number of these that will cover all the positive instances of the outcome.

5-5. fsQCA presents three solutions to each truth table analysis: (1) a “complex” solution that avoids using any counterfactual cases (rows without cases—“remainders”); (2) a “parsimonious” solution, which permits the use of any remainder that will yield simpler (or fewer) recipes; and (3) an “intermediate” solution, which uses only the remainders that survive counterfactual analysis based on theoretical and substantive knowledge (which is input by the user). Generally, intermediate solutions are best. For the truth table just presented, the three solutions are:

Complex:	severeaus*~govcorupt*pricerise*~notrepresv + priormob*severeaus*govcorupt*notrepresv + severeaus*govcorupt*pricerise*notrepresv
Parsimonious:	govcorupt*notrepresv + severeaus*pricerise
Intermediate:	severeaus*ricerise + notrepresv*govcorupt*severeaus*priormob

(Multiplication indicates set intersection—combined conditions; addition indicates set union—alternate combinations.)

Because they are logical statements, these two recipes for IMF protest can be factored. For example, the intermediate solution can be factored to show that severe austerity is present in both:

severeaus*(pricerise + notrepresv*govcorupt*priormob)

The expression indicates that IMF protest erupts when severe austerity is combined with either (1) rapid price increases or (2) the combination of prior mobilization, government corruption, and non-repressive regime.

Recall that severe austerity was close to perfection as a necessary condition in the crosstab analysis, so its importance here is not surprising.

Notice, however, that severe austerity is absent from one of the recipes in the parsimonious solution, suggesting that the parsimonious solution is probably too parsimonious. In other words, the parsimonious solution incorporates remainder rows that would not survive thoughtful counterfactual analysis.

Step 6: Evaluate the Results

6-1. Interpret the results as causal recipes. Do the combinations make sense? What causal mechanisms do they imply or entail? How well do they relate to existing theory? Do they challenge or refine existing theory?

6-2. Identify the cases that conform to each causal recipe. Often some cases will conform to more than one recipe and sometimes there are more cases that combine two (or more) recipes than there are "pure" instances. Do the recipes group cases in a meaningful way? Do the groupings reveal aspects of cases that had not been considered before?

6-3. Conduct additional case-level analysis with an eye toward the mechanisms implied in each recipe. Causal processes can be studied only at the case level, so it is important to evaluate them at that level.

The real test of any QCA result is how well it connects to cases. In this hypothetical application of QCA, there are no cases to connect to. Still, it is worth noting that the analysis started out as an examination of a single recipe, but ended up with two, each with different implied mechanisms. The mechanisms implied by "severe austerity combined with rapidly rising prices" are substantially different from those implied by "severe austerity combined with prior mobilization, government corruption, and a nonrepressive regime."

Step 7: Return to step 3 and repeat the analysis for the negation of the outcome.

Unlike statistical analysis, set theoretic analysis is asymmetrical. In fact, it is sometimes the case that the causal conditions that you have identified may work better at explaining the absence of the outcome than its presence. Consider the following truth table:

Advice (A)	**Commitment (C)**	**Shadow (S)**	**Inconvenient (I)**	**Reverberations (R)**	**Success (Y)**
1	0	1	1	1	1
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	1	1
1	1	1	1	0	0
1	1	1	0	0	1
1	0	0	0	0	1

"Advice" is a shared antecedent for "Success" but has no set-theoretic connection with "not-Success." "Inconvenient" is a shared antecedent for "not-Success" but has no set-theoretic connection with "Success."

Crisp versus fuzzy sets

Most introductions to QCA focus on crisp sets, just as this one does. Crisp sets are simple and straightforward and thus easy to present. However, social scientists (1) are often interested in phenomena that vary by level or degree (e.g., degree of membership in the set of democratic countries), and (2) dislike the sometimes arbitrary nature of dichotomization.

Fortunately, the procedures described here can be duplicated using fuzzy sets, which allow membership scores to vary from 0.0 to 1.0. Additional issues involved in the use of fuzzy sets include:

1. Calibrating the degree of membership in sets

2. Calculating degree of membership in a configuration, conceived as an intersection of fuzzy sets

3. Analyzing fuzzy subset relations

4. Constructing a crisp truth table summarizing the results of the fuzzy set analyses