Exercises for Set-Theoretic Analysis:
Online Appendix to "Set-Theoretic Methods for the Social Sciences"

Carsten Q. Schneider and Claudius Wagemann
with collaboration of Juraj Medzihorsky

Version 1.0

0. Introduction

0.1 What distinguishes set-theoretic methods from non-set-theoretic methods?

0.2 Please name different set-theoretic approaches.

0.3 What distinguishes Qualitative Comparative Analysis (QCA) from other set-theoretic methods?

0.4 What is the difference between QCA as an approach and QCA as a technique?

0.5 Which variants of QCA exist? What are their main similarities and differences?

1. Sets, Set Membership and Calibration

1.1 What is “calibration”?

1.2 What needs to be taken into account when turning raw data into set membership scores (i.e., when calibrating sets)?

1.3 What are the pros and cons of crisp set conditions and outcomes?

1.4 Discuss what is meant by difference in kind and difference in degree expressed by fuzzy set membership scores.

1.5 What is the difference between fuzzy set membership scores and probabilities

1.6 Calibrate the set “positive human development”, with the help of the ‘Human Development Index’ (HDI), available online. Use different calibration strategies (e.g., theoretical calibration, direct method, etc.).

1.7 In his famous article, Lipset (1959) analyzed the conditions for the outcome ‘survival of democracy’. The data file (‘Lipset_raw.csv’) contains the raw data of the relevant conditions, namely the set of developed countries, urbanized countries, industrialized countries, literate countries and politically unstable countries. Take the raw data contained in the file ‘Lipset_raw.csv’ and use the software fsQCA 2.5 in order to apply the direct method of calibration. For the condition DEVELOPED use 900 as the 1-anchor, 550 as the 0.5-anchor, and 400 as the 0-anchor. The three anchors are: for URBAN: 65; 50; 40, for LITERATE: 90; 75; 60, for INDUSTRIAL: 40; 30; 15, for UNSTABLE: 4; 10; 14 (the fuzzy set scores have been reversed so that they, indicate *stability*. This should be reflected in the name of the condition), for SURVIVED: 9; 0; -7

1.8 Take the same raw data of the file ‘Lipset_raw.csv’. Load them into Tosmana and set the threshold at 550 for DEVELOPED, at 50 for URBAN, at 75 for LITERATE, at 30 for INDUSTRIAL, at 10 for UNSTABLE and at 0 for SURVIVED.

2. Notions and Operations in Set-Theory

2.1 How is the fuzzy set membership of a case in an intersection of various sets calculated?

2.2 How is the fuzzy set membership of a case in a union of various sets calculated?

2.3 What is the universal set?

2.4 What does the 'Rule of the Excluded Middle' mean? Why can it not be applied to fuzzy sets?

2.5 Perform Boolean multiplication for the following pairs of expressions and simplify the result:

2.5.a) Expression A + B and expression A * B

2.5.b) Expression A + B and expression ~A + ~B

2.5.c) Expression A + B and expression A + B + C

2.5.d) Expression A * B and expression ~A * ~B

2.5.e) Expression A * (B+C) and expression B * (~A+C)

2.5.f) Expression (A + B)*(~A + C) and expression (A + ~B)*(~A + ~C)

2.6 Perform Boolean addition for the following pairs of expressions:

2.6.a) Expression A + B and expression A * B

2.6.b) Expression A + B and expression ~A + ~B

2.6.c) Expression A + B and expression A + B + C

2.6.d) Expression A * B and expression ~A * ~B

2.6.e) Expression A * (B+C) and expression B * (~A+C)

2.6.f) Expression (A + B)*(~A + C) and expression (A + ~B)*(~A + ~C)

2.7 Perform DeMorgan's Law for the following expressions and simplify, if possible:

2.7.a) A*B

2.7.b) ~A + ~B

2.7.c) (A + B) * (~A + (A*B))

2.7.d) (A + B + C + D) + (A*B*C*D)

2.7.e) A * (B * (C * (D+E)))

2.8 A case has a membership in set A of 1, in set B of 0, in set C of 0.6, in set D of 0.9 and in set E of 0.1. Calculate its membership in the following sets.

2.8.a) ~E

2.8.b) A*B

2.8.c) A*C

2.8.d) A*B*C

2.8.e) A + B

2.8.f) B + D

2.8.g) A + B + D

2.8.h) A + B + C + D + E

2.8.i) A*B*C*D*E

2.8.j) ~A + ~B

2.8.k) (A*B) + (C*~D)

2.8.l) (A+B) * (C+~D)

2.8.m) ~(A + B + C + D + E)

2.8.n) ~(A*~B + ~C + D*E)

2.8.o) ~(A*~B + ~C) + (D*E)

2.8.p) ~((A*~B + ~C) + (D*E))

2.9 Use TOSMANA to rerun the analysis of exercise 5

2.10 Use TOSMANA to rerun the analysis of exercise 7

2.11 Use the fsQCA 2.5 software (or EXCEL) to rerun the calculations of exercise 8.

2.12 Open the file 'Lipset_fuzzy.csv' and add, with the help of the fsQCA 2.5 software, the following new columns to the data sheet:

2.12.a) a column 'dandu' for the conjunction of 'developedfz' and 'urbanfz'

2.12.b) a column 'doru' for the disjunction of 'developedfz' and 'urbanfz'

2.12.c) a column ‘notd’ for the negation of ‘developedfz’

3. Set Relations

3.1 What do ‘equifinality’ and ‘conjunctural causation’ mean? Give an example.

3.2 How are the concepts of ‘sufficiency’ and ‘equifinality’ related to equifinality and conjunctural causation?

3.3 What is an INUS condition?

3.4 Describe the meanings of the following logical statements

3.4.a) $A \rightarrow Y$

3.4.b) $A + {\sim}B \rightarrow Y$

3.4.c) $A{\sim}C + {\sim}B \rightarrow Y$

3.5 Describe the meaning of the following logical statement and present it in a Venn diagram.

$A \rightarrow B$

where A denotes countries that have the €uro as their currency; B denotes membership in the European Union.

3.6 Which cells of the following cross tables do you need in order to analyze whether or not X is a sufficient condition for Y? Which ones do you need to analyze whether or not X is a necessary condition for Y?

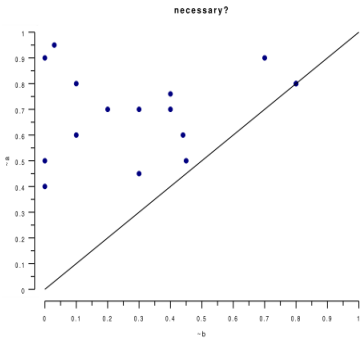
	~Y	Y
~X	(a)	(b)
X	(c)	(d)

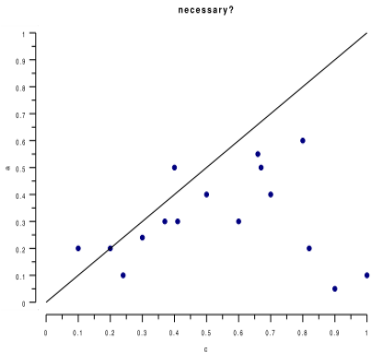
3.7 For each of the following two-by-two tables, specify for condition Z whether it is necessary, sufficient, both or nothing for the outcome ~K. (Note: one of the tables depicts outcome K!).

Z	1	Cases	-
	0	Cases	Cases
		1	0
		K	

Z	0	Cases	Cases
	1	-	Cases
		1	0
		~K	

3.8 Look at the XY plots: Which of the two indicates necessity of X for Y?





3.9 Open the file ‘Lipset_cs.csv’ and produce a crosstab between the outcome ‘survived’ and the condition ‘developed’. What kind of set relation do you see?

3.10 Open the file ‘Lipset_fs.csv’. Produce an XY plot with ‘survived’ being displayed on the Y axis and ‘developed’ on the X axis. Use ‘country’ as case ID variable. Which country violates the statement of sufficiency in the strongest way?

3.11 Open the file ‘Vis.csv’.

3.11.a) Produce an XY plot for U being displayed on the Y axis and PS~R being displayed on the X axis. (Attention: You first have to compute the combination PS~R).

3.11.b) Use the subset-superset analysis to find out the set relations between Y and PS~R.

4. Truth Tables

4.1 In fuzzy set analysis, a truth table is produced based on fuzzy data: what does it mean when we say that a given truth table row contains x number of cases?

4.2 Please construct a truth table showing the following conditions relevant to the electoral success of a political party.

Conditions:

I: Voter's Party Identification (voter identifies with the party/voter does not identify with the party)

A: Voter's agreement with the party's political statements (voter agrees with the statements/voter does not agree with the statements)

C: Voter's opinion of the political candidates (voter thinks highly of the political candidates/voter does not think highly of the political candidates).

Outcome:

Y: electoral success (success/no success)

5. Parameters of Fit

5.1 What is consistency and why is it an important concept?

5.2 What is coverage? How does the formula for coverage relate to the formula for consistency?

5.3 Provide the formula for the coverage of a necessary condition and describe what it means if the value of this formula is low.

5.4 Why does it not make sense to compute coverage before consistency?

5.5 In an analysis of sufficient conditions, we produce an XY plot with two inconsistent cases:
Case A has a membership value of 0.9 in the condition and 0.8 in the outcome. Case B has a membership value of 0.7 in the condition and a set membership of 0.1 in the outcome. Which of the two is more of a problem and why?

5.6 Imagine you have the following truth table. Do you see any problems you may have to solve before you can continue with your analysis? If yes, what can do about them, from the perspective of 'QCA as an approach'?

Row	Conditions			Outcome	
	High Public support	Many veto points	Strong business lobby	Strict climate policy (cases)	No strict climate policy (cases)
1	1	1	1	4	0
2	1	1	0	2	0
3	1	0	1	8	2
4	1	0	0	8	0
5	0	1	1	0	3
6	0	1	0	0	2
7	0	0	1	0	4
8	0	0	0	2	0

5.7 The following truth table, based on crisp-set data, shows in the columns Y and ~Y how often Y and ~Y are implied by the given configuration of conditions:

Row	A	B	C	Y	~Y
1	1	0	1	4	0
2	0	1	0	4	0
3	0	1	1	1	0
4	1	1	1	28	1
5	1	1	0	31	2
6	1	0	0	4	32
7	0	0	1	0	3
8	0	0	0	0	20

5.7.a) Calculate consistency of each row as a sufficient condition for Y.

5.7.b) Imposing a threshold of 0.9, logically minimize the truth table.

5.7.c) Calculate the solution coverage and the raw coverage of the single paths.

5.7.d) Why do we have to re-calculate consistency despite the fact that we have imposed a consistency threshold (0.9) for each truth table row?

5.7.e) Calculate the solution consistency and the consistency of the single paths.

5.8 Use the ‘Vis.csv’ data in order to analyze whether P, S and/or R are necessary conditions for U. Pay special attention to the coverage values.

6. Limited Diversity and Logical Remainders

6.1 What are logical remainders?

6.2 Mention reasons why limited diversity can occur in a dataset.

6.3 What are simplifying assumptions?

6.4 What is the most parsimonious solution term?

6.5 What are easy counterfactuals?

6.6 What are incoherent counterfactuals?

6.7 For the following truth table, derive the most parsimonious solution, the conservative solution and the intermediate solution (assumption: $A \rightarrow Y$). Perform the analysis first by hand, then with the computer.

A	B	C	Y
0	0	0	1
0	0	1	0
0	1	0	1
1	1	0	1
1	1	1	0

6.8 For the example in section 6.4.3.2, the directional expectations are made that each single condition (A-E) is expected to contribute to the outcome Y when it is present rather than absent.

6.8.a) What is the intermediate solution if only A is expected to contribute to the outcome Y and no assumptions is made about the other conditions (B-E)?

6.8.b) Which logical remainders have been included into the minimization in order to achieve that result?

6.9 Use the ‘Vis.csv’ data. How many logical remainders are there? Are they arithmetic, implausible or clustered remainders?

7. The Truth Table Algorithm

7.1 Which characteristics of your research design should you take into account when specifying the consistency threshold for a truth table row?

7.2 During the fuzzy set truth table algorithm, a truth table is produced which looks suspiciously similar to a dichotomous crisp-set data matrix. Why does a fuzzy-set analysis nevertheless produce a different result than a crisp-set analysis with dichotomized fuzzy values?

7.3 Consider the following raw data table, where A, B, C are the conditions and Y the outcome.

idcase	**A**	**B**	**C**	**Y**
1	1	0	0	0
2	1	0	0	0
3	1	1	0	1
4	0	1	0	1
5	0	0	0	1
6	1	0	1	0
7	0	0	1	0
8	1	0	1	0
9	1	0	1	0

7.3.a) Provide the truth table and add the column with the number of cases observed for each configuration.

7.3.b) Can you find some contradictory configurations? Which configurations?

7.3.c) Can you find some logical remainders? Which configurations?

7.3.d) Provide the conservative sufficient solution for this table.

7.4 Take the 'Vis.csv' data.

7.4.a) Find sufficient conditions for U. (Most parsimonious, conservative and intermediate solution. Choose a consistency threshold of 0.9. Use the directional expectations S $\rightarrow$ U and P $\rightarrow$ U). Interpret the role of S.

7.4.b) Find necessary conditions for U. (Test all conditions and their complements). Re-interpret the role of S.

7.4.c) Find sufficient conditions for ~U. (Most parsimonious, conservative and intermediate solution. Choose a consistency threshold of 0.8. Use the directional expectations ~S $\rightarrow$ ~U and ~R $\rightarrow$ ~U).

7.4.d) Which simplifying assumptions have been made for producing the most parsimonious and the intermediate solution for ~U, respectively?

7.4.e) Find necessary conditions for ~U.

8. Potential pitfalls in the Standard Analysis procedure and suggestions for improvement

8.1 Compare Standard Analysis (SA), Enhanced Standard Analysis (ESA) and Theory-Guided Enhanced Standard Analysis (TESA).

8.1.a) In which one(s) are only tenable assumptions used, but no untenable ones?

8.1.b) In which one(s) are only simplifying assumptions used, but no assumptions which do not contribute to parsimony?

8.2 This exercise is based on the Redding and Viterna data.

Outcome:

Success of left-libertarian parties in Western democracies [success] (SUCCES)

Conditions:

Rich society [pnb] (R)

High security spending [socsec] (S)

Corporatism [corpor] (C)

Left party in government [left] (L)

Proportional representational system [propor] (P)

8.2.a) Make an analysis of necessary conditions for the outcome. Consider C and L to be 'functional equivalents'.

8.2.b) Make an analysis of sufficient conditions for the outcome. Derive a coherent and plausible "enhanced most parsimonious solution".

9. Potential pitfalls in the analysis of necessity and sufficiency and suggestions for avoiding them

9.1 Which risks are run if necessary conditions are derived from an analysis of sufficient conditions by 'factoring out' those conditions which are contained in all paths of the sufficiency analysis? How can this be remedied?

9.2 Why is the PRI measure useful?

9.3 Which are the two sources of trivialness of a necessary condition?

9.4 This exercise is based on the Emmenegger data.

9.4.a) Make an analysis of necessary conditions for the *non-occurrence* of the outcome (without worrying about any 'functional equivalents').

9.4.b) Make an analysis of sufficient conditions for the *non-occurrence* of the outcome (most parsimonious, conservative and intermediate solution). When deriving the intermediate solution, make the following assumptions: V contributes to ~JSR rather in its presence than in its absence. All the other conditions contribute to ~JSR rather in their absence than in their presence.

9.4.c) What do the conservative and/or intermediate solution tell you about necessary conditions? What about ~P being a necessary condition?

10. Variants of QCA

10.1 In the two-step approach of QCA, a difference is made between ‘proximate’ and ‘remote’ conditions. How are they defined?

10.2 In a Multi-Value QCA (mvQCA), conditions A and B take on three values (0, 1, 2) and condition C four values (0, 1, 2, 3). Minimize the following expressions:

10.2.a) A{0}B{0}C{2} + A{0}B{1}C{2} + A{0}B{2}C{2}

10.2.b) A{2}B{1}C{0} + A{2}B{1}C{1} + A{2}B{1}C{2} + A{2}B{1}C{3} + A{1}B{1}C{3}

10.2.c) A{0}B{0}C{0} + A{0}B{0}C{1} + A{0}B{0}C{2}

10.3 Which version of QCA performs better when it comes to limited diversity: an mvQCA or a csQCA in which the multinomial categories of every condition are transformed into ‘dummy’ conditions?

10.4 Minimize the following expression of a temporal QCA (tQCA): A/B/C + A/B/~C + B/C/A + B/~C/A

10.5 How can the software package fsQCA 2.5 be used in order to deal with temporal processes?

11. Data analysis technique meets set-theoretic approach

11.1 What would be the effect if the membership value of a case was changed from below 0.5 to above 0.5 in a given condition?

11.2 Before starting a QCA, a researcher comes up with the idea that A or the combination of factors B and C are sufficient conditions for the outcome (A + BC $\rightarrow$ Y). Her results reveal the following solution: A~B + C $\rightarrow$ Y. In which sense do her empirical finding support or contradict the prior theoretical expectations? (For simplicity reasons, do not worry about consistency values).

11.3 Cases with a membership value of < 0.5 in the condition and of > 0.5 in the outcome are perfectly consistent cases with regard to sufficiency.

11.3.a) However, what is the problem?

11.3.b) Why is it logically flawed to compare such a case with a consistent case with a value of > 0.5 both in the condition and in the outcome ("typical case")?