To cite this annex, use the Creative Commons method:
Olsen, Wendy (2021), *Systematic Mixed Methods Research for Social Scientists Annex of Practical Techniques*, Creative commons, URL https://github.com/WendyOlsen/SystematicMixedMethods

Systematic Mixed Methods Research for Social Scientists

By Wendy Olsen

2021

Online Annex

This annex supplements the book **Systematic Mixed Methods Research for Social Scientists** by Wendy Olsen (London: Palgrave, 2021). The book is about merging the interpretation of mixtures of data from quantitative and other methods. This annex contains practical and instructional material for some of the activities. By offering some details that arose during face-to-face teaching of mixed-methods research design, my aim is to make it more accessible to approach the tasks from a mixed-methods vantage point.

Acknowledgements: I gratefully thank the Cathie Marsh Institute and Methods@Manchester at the University of Manchester. They hosted summer schools, 'What is…' webinars, and one-day training events. The people in these organisations were helpful in promoting mixed methods. The other acknowledgements to funders are located in the book itself.

Annex Contents:

Introduction

The material in this Annex is mainly related to Qualitative Comparative Analysis, ie Chapters 6-7 of the book, and qualitative interpretation, ie Chapter 8. In both areas I reviewed some innovations. In regard to QCA I proposed a distance measure to gauge goodness of fit of a pattern to a hypothesis, and I provided a freeware package, released via the Github free online repository, to allow quick visual examination of all the scattergrams implied by one truth table. This is useful for both crisp and fuzzy-set analyses. The Github package, called 'fsgof' (fuzzy-set goodness-of-fit) has undergone one version change so far. Comments are welcome.

In regard to qualitative interpretation I surveyed a range of statistical methods combined with qualitative coding and discourse-analysis to make it possible to easily analyse large corpuses of textual data. I argued that methods centred on discourses and meanings are different from 'data mining' using the following distinction. If you were data mining you might have routines to follow, which a computer can implement, and thus the wider scope of studying a larger corpus would be handled mainly by artificial intelligence. In other words a black-box mechanism would lead to newly transformed data summaries, which you can then interpret. On the other hand the keyness analysis suggested here illustrates a different kind of method (and it is just one illustration of the key point): It requires and acknowledges our human input at three critical points. First in deciding on what data to use, within a particular context and given that knowledge about that scope and time will be gathered from multiple sources. Secondly, the researcher decides what kinds of phrases will be coded (e.g. their meanings analysed; single words versus other phrase sets such as sentences; words of what length; or images or video. Thirdly, the researchers will again intervene at the stage of deciding how to group the words, phrases, or meanings into discourse topics and how to break up the topics into discrete discourses. Finally both methods have a concluding interpretation stage. These stages involve considerable synthesis of prior knowledge, which you could have from fieldwork, from reading ethnography or other qualitative reports, from scientific literature, and from your personal experiences. Thus, the methods I propose in this Annex are not the same as data mining whilst most of them could be used by data miners.

I discussed in Chapter 8 that the notion of what 'expertise' is needed to be a scientist can include knowledge about macro and meso objects, not just the micro objects that were 'observed' or about which data have been provided. Thus, we discern linkages between history and current events, and researchers raise awareness of large-scale social trajectories… Overall the scientific realm is not demarcated by choice of a particular method, nor merely by using evidence, nor by internal or external validity alone, but instead by a range of 'tenets' of epistemology (Chapter 9). These tenets are not presented in this Annex which is much more procedural. The book, instead of just being on methods, is methodological.

In brief the innovations in this Annex relate to social science and not to mathematics, although as usual I have used mathematics at key junctures.

I welcome feedback. The copyright status of these online annex materials should be respected by careful citation.

Conducting Qualitative Comparative Analysis (QCA).

Chapter 6

START BOX ANNEX 6.1:

6.1 The Technical Protocol for Using fsQCA Freeware

Step 1: Run the Software

When you want to run QCA you may use either a simple spreadsheet with your own formulas, or fsQCA freeware. See details in Ragin, Charles, and Sean Davey (2009, 2014). Their team's fs/QCA [a computer programme], is in version 3.[1] First save the v3.0 fsQCA zip file to C: drive. Then unzip the files via 'extract all files'. Next run the file fsqca.exe. fsQCA is a relatively small program. If you prefer to use a spreadsheet, STATA, or R, look up the relevant instructions.

- For R, use Thiem and Dusa, 2013 as instructions; load qca or qcapro;
- For STATA, use Longest and Vaisey, 2008, as instructions; scc install fuzzy;
- For spreadsheets, you can do the truth table and consistency values for a complex project using formulas as is demonstrated in the files held at the JISCMAIL repository for QUAL-Compare see https://www.jiscmail.ac.uk/cgi-bin/filearea.cgi?LMGT1=QUAL-COMPARE, for which registration is at https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=QUAL-COMPARE, accessed 2020; and
- For an example of Python programming see McLoughlin and Olsen's fsgof programme in Github, which is open source.

The spreadsheet method is one which every aspiring QCA user may want to try, because it helps consolidate our awareness of the links between consistency in its two forms: necessity

[1] Ragin, Charles, and Sean Davey (2014) *fs/QCA [Computer Programme]*, Version [3.0]. Irvine, CA: University of California. The instructions for loading are at: http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml accessed 2020. Users will prefer Version 3.0 or later, and must not load Version 2.0 (held at http://www.u.arizona.edu/~cragin/fsQCA/software.shtml, and dated 2009) unless it is for scholarly reasons of identifying which distinct formulas and methods of fsQCA version 2 versus version 3 gave particular results.

and sufficiency (see also Chapter 6 of the present book). A close look at the file shown below may be helpful (see also URL url https://www.jiscmail.ac.uk/cgi-bin/filearea.cgi?LMGT1=QUAL-COMPARE&X=8FCAA2709D50A7369C&Y=wendy.olsen%40manchester.ac.uk&a=get&f=/NotesonByrneTableRawMaterialAnnexTable.xlsx, accessed 2020, which is discussed below).

For a quick start, grab the data set from one of the exemplars held in the Bibliographic reference area of www.compasss.org. These datasets are also available in the best journals as Supplementary Files.

Step 2: Check Your Data for a Good Spread

When applying QCA or any statistical test of causality over a domain, which in fuzzy set analysis is from 0 to 1, you must have a sample that has some representation at different parts of the X domain and the Y domain. In QCA therefore, look at your case plots for each configuration. In the Appendix, I showed that our fsgof software – which stands for Fuzzy Set Goodness of Fit software - can quickly illustrate all your configurations for you.

Note: If the sample has failed to give a good range of X, then you cannot test that X for causality.

Step 3: Check for Necessary Causes, and Remove them; Write Them Up

Using fsqca, measure the 'consistency' of each configuration with the necessary pattern, which is when Y is included in X. That is, all cases which are high on Y are high on X, and the exceptions are not enough to reduce the 'consistency' value below the cutoff you choose, likely to be 0.75 or 0.80.

Step 3: Examine Causal Pathways for Sufficiency; Write Them Up

Using fsqca, measure the 'consistency' of each configuration with the sufficiency pattern. In this pattern, X is included in Y, nearly all cases have a Y value exceeding the X value, and therefore whenever X rises, Y is at least as high, or higher. In other words, all cases which are high on X are high on Y, whilst there could be cases of Y which occur for other reasons, when X is low.

[START ANNEX BOX 6.2]

Key Terminology for Using CMO with QCA

Context-Mechanism-Outcome (CMO) is an approach developed by Pawson and Tilley (1997) which stresses a realist approach helps in evaluating what 'works' to generate a good outcome. They analysed how QCA logic can supplement the standard randomised-control trial logic. They showed that a mechanism X may not work without its standard, often taken-for-granted context. The contextual factors A, B, and C then are required parts of a configuration that includes X to cause Y. We might write:

AX + BX + CX → Y where + represents 'OR' and → represents sufficient cause.

In this statement, A, B, and C are parts of the three pathways to Y, while X now appears as a necessary cause. If the situation were different, then no single element would be 'necessary for Y', as shown below:

AX + BX + C → Y where + represents 'OR' and → represents sufficient cause.

Which is equivalent to: (A + B)*X + C → Y. (Read this as A or B combined with X, or C alone, are sufficient for Y.)

Distance: In general distance between points is measured using 'Euclidean' measures, such as the triangular measure of distance based on Pythagoras' theorem. (The square root of the sum of squared lengths A and B will be the length C that completes the right triangle.) In Boolean data however, distance may be measured in other ways because the algebra rules chosen are Boolean rather than Euclidean. In the fuzzy set analysis that I proposed, the fuzzy sets were adapted back into ordinary Z scores as advised by

Equifinality: Another word for multiple sufficient pathways of cause from a set of explanatory factors, represented by the elements of the vector X, to the outcome Y.

Quasi-necessity: A type of causality in which X is almost always part of the causal conditions for an outcome Y, and X is thus necessary for Y. The degree of consistency should be above the threshold level of 0.75 or 0.80 as described in several textbooks.

Quasi-sufficiency: A type of causality in which some configuration (which might be PQ or ABC or PB, or might involve the absence of a factor, such as B, and thus written Pb) is enough in itself to cause, or result in, an outcome Y. The list of elements such as PQ can be summarised as X where X is a vector of values, one for each case, which have both P and Q. Thus X would hold the values {1,1} for all instances of PQ but it is an empirical question whether PQ is sufficient for Y. We thus examine the data, for example as shown below.

Table A6.1: **Illustrative Crisp Set Dataset**

CaseID	P	Q	Y
1	1	1	0
2	1	1	1
3	1	1	1
4	1	1	0
5	1	1	0
6	1	1	1
7	1	1	1
8	1	1	0

It will not be crucial to know what happened to Y when PQ were absent, but it will affect the 'consistency' of the measure of sufficiency. Therefore we also look at cases without PQ:

Table A6.2: **Continuation of Illustrative Dataset**

CaseID	P	Q	Y
9	0	1	0
10	0	1	0
11	0	1	0
12	0	1	0
13	1	0	0
14	1	0	0
15	0	0	0
16	0	0	0

As shown here, the cases without PQ do not have Y; and many of the cases with PQ do have Y. However, consistency will be low, as there are cases with PQ which do not have Y. We therefore move to the 'quasi-sufficiency' concept. In truth-table terms, we can summarise the situation as shown below.

Table A6.3: **Sample Truth Table Derived From the Above Data**

P	Q	Y	N
1	1	1	4
1	0	1	0
0	1	1	0
0	0	1	0
1	1	0	4
1	0	0	2
0	1	0	4
0	0	0	2
Total N:			16

Sum of the Intersection of PQY: 4

Sum of the Y: 4

Consistency for sufficiency of PQ for Y: 1.

In the above example, seeking the sufficient conditions for an outcome Y, PQ is the joint element that I am going to call X. X is shown empirically to be sufficient for Y. The degree of consistency should be above the threshold level of 0.75 or 0.80 as described elsewhere. We usually call PQ a configuration, and the % of Y which were Y=1 for PQ is 100%. However, we used a formula which could consider the other case possibilities, Pq, pQ, and pq, which are all part of the configurational approach. The consistency formula uses Boolean algebra not Euclidean algebra.

When using fuzzy sets, the concept of a percentage should not be used.

Remainder: a configuration, of which there are no cases.

END of Box 6.2**

[END BOX]

To cite this annex, use the Creative Commons method:
Olsen, Wendy (2021), *Systematic Mixed Methods Research for Social Scientists Annex of Practical Techniques*, Creative commons, URL https://github.com/WendyOlsen/SystematicMixedMethods

[START BOX ANNEX 6.3]

Key Terminology for Fuzzy Set QCA:

Configuration: a set of circumstances.

Fuzzy set intersection: a measure, ranging from 0 to 1, which represents the measurement score of a given case's membership in a set X where X is the intersection (that's the shared part) of some other sets such as X1, X2, X3 and X4. We write the intersection as X1∩X2∩X3∩X4=X. Then if then we want to intersect this X with Y, we could write X∩Y and in algebra, we also can work it out as Min(X, Y).

Union: A mixing of two sets in which their maximum membership score is given as the result. In Venn diagram terms this 'union' might look like the combination of both sets, inclusively. We sometimes call it the inclusive or, denoted ∪. Another symbol for 'union' is +. Thus, X or Y can be written as X+Y and is also written in a spreadsheet formula as Max(X, Y).

Fuzzy Sets: a set membership score assigned to a bunch of cases, or a single case; Each fuzzy set is a measure, ranging from 0 to 1, which represents the measurement score of a given case's membership in a set X. For 'X' we can substitute a long qualitative description of a particular feature, such as marriage by common law or marriage by cohabitation.

Mechanism: a cause. A thing which can produce an outcome. This 'thing' is of a certain type, and its nature is described by a key characteristic. This characteristic may be used as the name of the fuzzy set that represents the mechanism, e.g. Literacy.

Necessity: A type of causality in which X is always part of the causal conditions for an outcome Y, and X is thus necessary for Y.

- instance: a positive or true value for the outcome variable

Pluralism: Taking into account two or more disciplines when studying a social or natural/physical problem, or doing research that is multidisciplinary.

Case: a situation which has a set of circumstances, which are being compared across a bunch of cases, in each of which the outcome may be positive/negative, true or false, absent or present. When we name the cases, we often impose a condition that they be of the same 'kind', having some essential features in common. However, cases also have characteristics that are diverse. We measure these diverse ones as variates.

Cause and effect: An enumeration of two factors, one having the capacity or tendency to create the conditions for the occurrence or existence of the other. The cause may pre-exist or co-exist with the thing that we consider as its effect.

Comparative method: A research approach that compares fundamentally distinct yet comparable cases, perhaps internationally across countries or regions, or across time of space (comparing wars; comparing riots and the reasons for riots; and so on). In general the comparative method can apply many other mixed-methods approaches but tends to be ambitious in drawing out similarities and differences.

Consistency: a measure ranging from 0 to 1 indicating how closely a pattern of cases matches the hypothesis that X is sufficient for Y, given that the X and the Y are each either 'fuzzy sets' or 'crisp sets'. When consistency is 1, and we are seeking evidence for 'sufficiency', every case that has X will have Y but not necessarily vice-versa. Another way to say this is X is sufficient to cause Y.

Consistency for necessity: a measure ranging from 0 to 1 indicating how closely a pattern of cases matches the hypothesis that X is a necessary condition for Y to occur. 1 indicates all Y's have X, and therefore that X appears necessary for Y to occur.

Contradiction: a configuration in which some cases have a positive instance and others do not. (Contradictions occur only in crisp set situations.)

[END BOX ANNEX 6.3]

START BOX ANNEX 6.4

Table A6.4: The Coded Data of the Lam-Ostrom Exemplar (2010)

A	R	F	L	C	Success in W	Not W
0	1	0	1	1	1	1
0	1	1	1	1	2	0
1	1	0	1	1	2	0
1	1	1	1	1	2	0
0	0	0	0	1	1	0
0	1	0	0	1	1	0
0	0	1	1	1	0	1
1	0	0	0	0	0	1
1	1	1	0	0	1	0
1	1	0	0	1	1	0
1	1	1	0	1	1	0

A full set of equations representing these data are shown below. The equations offer a full account of the pathways sufficient to have caused good water supply, W, in the watersheds studied in Nepal.

There were 14 cases falling into 11 configurations. Recall that a configuration is a distinct set of explanatory circumstances. The notation is as follows. A condition's presence is indicated by capital letters, and its absence by small letters. Thus cases with ARC had A, R and C features. The ARC cases were 5 in total, and they all had success in the outcome W. Therefore, A R and C may play a role in the explanation of W. However the QCA method also looks at failures when assessing if A, R or C is sufficient for W. Here we can note two points:

First testing for necessity, ARC is not necessary for W. Neither is A, nor R, nor C independently taken.

Secondly we start writing out what pathways are sufficient for W.

These are shown in Equation 6.1, 6.1', 6.1'', and 6.1'''. The use of prime (denoted ') helps us keep in mind these are revised representations of the same dataset.

ARC or aflC or ARFl or RFLC => the W outcome (sufficient for good water supply)

Eq. 6.1

ARC + aflC + ARFl + RFLC => the outcome W. Eq. 6.1'

AR (C + Fl) + ... => W Eq. 6.1''

AR (C + Fl) + C (afl + RFL) => W Eq. 6.1'''

These steps use Boolean algebra. Here it is just used to re-present the data. Yet a representation of the data is also a 'finding'. The findings offer summary data. Based on these, we can write out an interpretation.

There was no factor that was 'necessary' for W overall. The consistency of that result is 0. This measure used is consistency of necessity.

The consistency of the solution shown in Eq. 6.1 is 1.0. The measure used here is the consistency of sufficiency. It shows that the crisp set results match perfectly the hypothesis that these four pathways are sufficient to explain the outcome W.

This study may appear small-scale to those with quantitative survey research experience. The cases do exhibit limited diversity. Limited diversity means that not all possible permutations actually occur. The response of realists to this situation is summarised by Thomann and Magetti (2017: 15): "Case-oriented QCA approaches may mitigate the potential problems of limited diversity for internal validity through the additional use of case knowledge. Comparative case studies can substitute for counterfactual reasoning, and single-case studies can help to support the causal interpretation of INUS conditions (Rohlfing and Schneider 2013, 2016; Schneider and Rolfing 2013, 2016)." INUS conditions are those which involve some necessary condition(s) as part of a set of explanatory factors that work together as a sufficient pathway. The N in INUS refers to 'necessary condition' and the S stands for sufficient condition. By contrast, 'necessary cause' in QCA tests refers to an explanatory factor which is apparently necessary across all the known cases.

See also the fuzzy set reasoning and testing held as supplementary material for Hinterleitner, et al., 2015, at URL https://ejpr.onlinelibrary.wiley.com/doi/full/10.1111/1475-6765.12142 accessed 2019.

END BOX

Table A6.5 shows how the elements of fuzzy and crisp set analysis fit together using a ratio approach to the two key elements of the 'consistency' measure – also known as the inclusion ratio – which is also treated further in Annex Box 7.1, further below.

Table A6.5: Illustration of Consistency and Coverage with Crisp Sets

Illustration of Consistency and Coverage with Crisp Sets
Based on Byrne's Chapter 14 in Byrne & Ragin, eds., 2009

From *The Handbook of Case-Based Methods*, pp 260-268, 2009 (London: Sage).

Accessed Aug. 2013						Number	% With	Sufficiency		Necessity		Def'n of	CONSISTENCY for SUFFICIENCY	COVERAGE for SUFFICIENCY	DENOMINATOR FOR BOTH,
X1	X2	X3	X4	X5	X6	of Cases	Y	Consistency	Coverage	Consistency	Coverage	X	NUMERATOR	NUMERATOR	
1	1	1	1	1	0	27	82	0.82	0.31	0.31	0.82	X1X2X3X4X5	22	27	71
0	1	1	1	1	0	18	61	0.61	0.15	0.15	0.61	X2X3X4X5	11	18	71
0	0	0	0	1	0	12	8	0.08	0.01	0.01	0.08	X5	1	12	71
0	1	1	0	1	0	12	17	0.17	0.03	0.03	0.17	X2X3X5	2	12	71
0	0	1	0	1	0	11	27	0.27	0.04	0.04	0.27	X3X5	3	11	71
1	1	1	1	1	1	9	78	0.78	0.10	0.10	0.78	X1X2X3X4X5X6	7	9	71
0	0	1	1	1	0	8	63	0.63	0.07	0.07	0.63	X3X4X5	5	8	71
0	1	0	0	1	0	7	0	0	0.00	0.00	0.00	X1X2X3X4X5	0	7	71
1	0	1	1	1	0	6	33	0.33	0.03	0.03	0.33	X1X2X3X4X5	2	6	71
0	1	0	1	1	0	4	25	0.25	0.01	0.01	0.25	X1X2X3X4X5	1	4	71
1	0	0	0	1	0	4	0	0	0.00	0.00	0.00	X1X2X3X4X5	0	4	71
1	1	0	1	0	1	1	100	1	0.01	0.01	1.00	X1X2X4X6	1	1	71
1	1	0	1	1	1	1	100	1	0.01	0.01	1.00	X1X2X4X5X6	1	1	71
1	1	1	0	1	0	1	100	1	0.01	0.01	1.00	X1X2X3X5	1	1	71
Total Overall						155	45.8	0.46	0.00	1.00	0.46	Any of above	71	155	71

N1+ N2 + N3+ N4

Proportion of This Config's Cases with Y=1

N2+N3

N3+ N4

=(SUm(N3)/ (SUM(N3+N4))

=(SUm(N3)/ (SUM(N2+N3))

Overall Sum(N2+ N3)

By Wendy Olsen 2013
Uploaded to Jisc Email list 'QUAL-COMPARE'
Open to all. This table can be used by anyone.

For a given line in the table above, **X** is the intersection of selected elements from{X1, X2, X3, X4, X5, and X6} = min(some of{ X1, X2, X3, X4, X5, X6}).

Y = 1	***n2***	***n3***
Y = 0	***n1***	***n4***
	X = 0	**X** = 1

By Wendy Olsen
2020

comments to: wendy.olsen@manchester.ac.uk

Notes to Table A6.5

The filename NotesonByrneTable.XLSX has three tabs and can be found at the jiscmail repository, url https://www.jiscmail.ac.uk/cgi-bin/filearea.cgi?LMGT1=QUAL-COMPARE&X=8FCAA2709D50A7369C&Y=wendy.olsen%40manchester.ac.uk&a=get&f=/NotesonByrneTableRawMaterialAnnexTable.xlsx, accessed 2020.

Note that whilst the front 'tab' of the spreadsheet is shown above, there are two more tabs. Tab 2 of the spreadsheet contains the truth table and tab 3 is the raw data. The labels on the columns are as shown in the original book chapter, x1sixth x2sen x3lowabs x4notdep x5coed and x6relig, respectively. These refer to the following variates:

Has a sixth form? Has high SEN rate? Has low absenteeism? Is not in a deprived area? Is it co-ed by sex? Religious school?

START BOX ANNEX 7.1

7.1: Using FSQCA Freeware for QCA and Fuzzy Set analysis Truth Table reduction

Chapter 7

*** ONLINE ANNEX MATERIAL

Calculating 'consistency' levels for all permutations can be carried out in a spreadsheet or using software. The 'consistency' level as Ragin calls it uses the same formula as the fuzzy-set inclusion ratio in mathematics (Smithson and Verkuilen, 2006; Caramani, 2009). It is sometimes called measuring the subset relationship of X and Y. The consistency level depends on two parts of a ratio: the fuzzy set intersection of X and Y, and the fuzzy set sum over all cases in a configuration for X. The specific ratio that is usually used to measure consistency of sufficiency is represented as Eq. 1:

Consistency $(Yi \leq Xi) = \sum(\min(Xi, Yi)) / \sum(Xi)$. Eq. 1

The 'sigma' is a summation to show that we aggregate by adding up fuzzy set intersection scores across the cases in the configuration X.

Rihoux and Ragin, 2008, define two such ratios - one for sufficient cause, and one for necessary cause.[2] For the study of sufficient causes, the minimum of {Xi, Yi} is how we measure the 'intersection' of X and Y. Recall that a fuzzy set intersection is an 'AND' operation. (Most people call this a Boolean operation, but it is actually a rule in fuzzy-set logic, see Bergmann, 2008. See also Mendel and Korjani, 2012, 2018 for the formulas.) This aggregation over cases, within one configuration, to get the sum of min{Xi, Yi}, allows us to calculate the proportion of cases in which the outcome, Yi, is lower than the causal variate Xi. The causal variate Xi is the baseline so its aggregate sum goes in the denominator.

The cases where Yi is "Lower" in ordinal, not cardinal, measurement terms are the cases that contradict they hypothesis that X is sufficient for Y. The consistency formula is Sum(Min(X, Y)) / Sum (X), where each element is a sum over all cases within the configuration that we are testing. This takes a value from 0 to 1, with 1 occurring if all cases that have X also have

[2] For ease of reading I am not covering either 'necessary consistency' or 'coverage' here. These are easy additional concepts because of the symmetry of the mathematics of consistency. See Ragin (2009) or Rihoux and Ragin (2008) for details.

Y and there is no case where Xi is higher than the corresponding Yi. In graphical terms, the triangle above the diagonal is full of cases, while the other triangle is empty.

If we want to measure the consistency of the X variate being **necessary** for the Y outcome, we are looking for the inverse subset relationship, where X always is less than Y. Therefore we want the ratio of Sum(Min(X, Y)) / Sum (Y). This takes a value from 0 to 1, with 1 occurring if all Y cases have X. In graphical terms, the triangle **below** the diagonal is full of cases, while the other triangle is empty (see also the Appendix to the book).

When the algebra is extended to the multivariate case (Rihoux and Ragin, 2008, Chapter 3:17), the numerator of consistency is visible in an X-Y plot as the sum of horizontal distances from the vertical axis to each point in the graph. For **sufficiency** to be strongly supported by the data, most points should lie in the upper left, not the lower right triangle. We are measuring how close the non-conforming cases (which lie in the lower right triangle) lie to the diagonal line, relative to the average.

QCA consistency values are non-symmetrical in Y. If we reverse the procedure and study not-Y, we would get different results.

X in these formulas might stand for several characteristics combined, e.g. AFR or AfR or AR. Thus X is itself a fuzzy-set intersection. This is why the formula is so powerful: it is used to test multiple causes (in their fuzzy-set intersection, which is the minimum of all the parts of the set, min(A and F and R)), not just to test single causes in isolation. It allows for overlapping causality. This is why QCA fits the Context-Mechanism-Outcome approach so well.

*** END OF BOX***

Testing Hypothesis About Sufficient Pathways with FSGOF

START BOX FOR ONLINE ANNEX 7.2

7.2: How to Use The Fuzzy-Set Goodness of Fit (fsgof) Testing Software

Re Chapters 6-7. See URL https://github.com/WendyOlsen/fsgof, accessed 2019.

The fsgof program was created by Wendy Olsen and John McLoughlin. It runs fuzzy set goodness of fit tests, providing results in tables and plots. The advantages of the programme are twofold:

(a) The fsgof program plots include easy-to-see configuration scattergrams for every extant X configuration.
(b) The results tabulates the Consistency of Sufficiency along with the sample size and F test for distance to the sufficiency triangle. The F test results are based on Z score equivalents to each fuzzy set, and they again cover every extant X configuration.

Advantages:

Firstly, if the QCA data are a random sample, then we can generalize to the population using *fsgof*. For example, a confidence interval for the consistency measure can ultimately be set up. So far we have worked up an F test of how the variance is affected by the hypothesis versus the sample data. Using the 'deviance measure', here denoted Dsuff, also known as an aggregate distance measure, we are able to bring in two kinds of variation across a population (Eliason and Stryker 2009). In future, using such methods, Bayesian methods of simulation which are not dependent upon normal distributions can allow relationships to be estimated with predictive intervals. A Bayesian predictive interval is nearly the same as a confidence interval in frequentist statistics (e.g. 95% interval). However, the Bayesian approach using Markov Chain Monte Carlo simulation modelling also allows measurement error to make us provide a more conservative interval than we otherwise would. Thus while in standard QCA we assume no measurement error, in mixed-methods Bayesian QCA we could allow for measurement error. The current package has not taken up measurement error but has allowed for random sampling error.

Secondly, the *fsgof* analysis package is partially automated using a new two-step online procedure. The user downloads the freeware then submits to it a spreadsheet of data. It allows up to four Outcomes (Y fuzzy sets). Data can be crisp or fuzzy, or a mixture. First *fsgof* produces the consistency measure, and second it gives the F test results for each permutation of the causal conditions.

Technical Requirements:

The computer will need Python Anaconda programmes on the c: drive. Make a directory to receive the output, specifically called c:\fsgof and then prepare your single-line command, which you will run to get the consistency values and the F tests. Download the whole package to get up-to-date instructions.

How to Use the Freeware:

The scatterplots are key results of the programme. They will go in the directory you created and from which you ran the command. The program produces a table of output using csv format and this is provided in Word format alongside the csv file.

On Windows machines use: START > COMMAND PROMPT then cd c:\fsgof.

For example, suppose I have put the Python Anaconda programmes on the F: drive. Then run the command like this:

```
C:\fsgof> f:\python.exe CDsuff.py c:\data\****.csv 4 > op4.doc
```

The command options are as follows:

Python inputfilename Ychoice ">" [that means output the text file to] textfileoutputfilename

A simple command line example reflecting the standard settings is:

```
C:\fsgof>  python.exe CDsuff.py c:\fsgof\inputIndia.csv 1 > outputforY1India.doc
```

NOTES:

```
# The developers are John McLoughlin and Wendy Olsen of the
# University of Manchester.
# This work is released under the Creative Commons Licence.
# You are free to use, change and distribute this work as long
# as you cite:
```

Fuzzy Set Goodness of Fit Tests Version 1 or Version 2.

JM & WO 2016/06/28 URL https://github.com/WendyOlsen/fsgof , accessed 2019 (Version 1) and for version 2, 2019, see URL https://github.com/luminous-badger/cdnec, accessed 2019.

This program implements Ronggui Huang's 2010 R program for

Goodness-of-Fit Tests and Descriptive Measures in Fuzzy-Set Analysis. See also: Eliason S. & Stryker R. 2009. *Sociological Methods & Research* 38:102-146.

1. Conducting a Qualitative Interpretation using Warranted Arguments

This second part of the online Annex offers details about gathering, summarising and coding the data from projects that have an interview element. The methods can be used on a variety of textual and image data, including multi-media sound and video files, where we can code the material.

The first Annex table offers a list of words exceeding 4 letters each, as suggested in the book's Chapter 8. The list has been organised according to alphabetical order within nine groups. The group headings at the top indicate nine main topics of the material. Using both Chapter 8 and the works cited there, you can work out how to carry out similar analyses to prepare a data summary of this kind on a corpus.

The word counts are shown in the columns marked as 'v2 and v4', meaning variables 2 and 4 from a Stata file. The variable 1 column contains the word text and this one is a 'string' variable, or character type variable. V2 is the word count of this word in the data corpus. V4 is the word count of the same word in the British Baby National Corpus. (The 'Baby' is much smaller than the main British National Corpus and was workable on our desktop PCs.) The odds of a word occurring is the word count divided by the remaining number of words in that corpus. (Remember that odds are n divided by total N – n.)

In the spreadsheet some rounding occurs, making the odds appear sometimes not to correspond correctly to the input data. Even if a word is rare, and we use tiny decimal numbers to represent its odds of appearing, the ratio of two odds is still interesting and is a viable measure. The odds ratio here is the odds of the word in v2 divided by the odds of the word in v4. The only situation where the odds ratio is not viable is when a word is not found in one of the corpuses. If it is not in the British Baby National Corpus we left the word out of this analysis.

START BOX 8.1

Units of Discourse and their Prevalence (Evidence from South India 2006-7)

[<<**To be provided separately. It is a full page. The file name in word is: AnnexTableWithGroupedWordsinTopicsIndia.docx.

It is a single page, csv format, table. Free nodes for a study with 39 interviews comprising 39 internal files.

I insert the file below FYI.**.]

To cite this annex, use the Creative Commons method:
Olsen, Wendy (2021), *Systematic Mixed Methods Research for Social Scientists Annex of Practical Techniques*, Creative commons, URL https://github.com/WendyOlsen/SystematicMixedMethods

feelings	agriculturemarriage	religion	money	behaviours	family	status	appearance	work	v1	v2	v4	oddsbnc	oddsssi	oddsratio
			bribe						bribe	2	16	0.0004	0.0008	2.0059
			bribery						bribery	2	13	0.0003	0.0008	2.4690
			creditors						creditors	2	10	0.0003	0.0008	3.2099
			income						income	78	246	0.0062	0.0325	5.2191
			indebted						indebted	1	5	0.0001	0.0004	3.2091
			indebtedness						indebtedness	1	11	0.0000	0.0004	16.0469
			money						money	555	1638	0.0430	0.2891	6.7171
			passbook						passbook	1	1	0.0000	0.0004	16.0469
			passing						passing	83	583	0.0149	0.0347	2.3282
			repaid						repaid	2	10	0.0003	0.0008	3.2099
			repay						repay	32	36	0.0009	0.0131	14.4322
			rupees						rupees	167	3	0.0001	0.0724	957.4764
			salary						salary	14	75	0.0019	0.0057	3.0056
			spent						spent	55	391	0.0099	0.0227	2.2849
altercation				angry					angry	35	160	0.0040	0.0143	3.5449
argument				argument					argument	108	635	0.0163	0.0456	2.8071
ashamed									ashamed	5	35	0.0009	0.0020	2.2942

disheartened				disheartened	1	2	0.0001	0.0004	8.0232
displeased	displeased			displeased	1	4	0.0001	0.0004	4.0114
displeasure	displeasure			displeasure	1	5	0.0001	0.0004	3.2091
entrapped			entrapped	entrapped	1	3	0.0001	0.0004	5.3487
felicitous				felicitous	1	2	0.0001	0.0004	8.0232
giddy				giddy	2	7	0.0002	0.0008	4.5860
gladly				gladly	3	2	0.0001	0.0012	24.0892
heartache				heartache	1	4	0.0001	0.0004	4.0114
helpless				helpless	25	39	0.0010	0.0102	10.3773
honor		honor		honor	9	3	0.0001	0.0036	48.2944
jealous				jealous	16	32	0.0008	0.0065	8.0661
jealousy				jealousy	5	32	0.0008	0.0020	2.5094
worthless		worthless		worthless	6	9	0.0002	0.0024	10.7174
	abuse			abuse	35	139	0.0035	0.0143	4.0826
	altercation			altercation	1	7	0.0002	0.0004	2.2921
	behave			behave	51	105	0.0027	0.0210	7.9341
	behavior			behavior	21	2	0.0001	0.0086	169.8613
	bickering			bickering	3	6	0.0002	0.0012	8.0289
	impertinent			impertinent	1	7	0.0002	0.0004	2.2921

nicely		nicely	15	65	0.0016	0.0061	3.7182
obstinacy		obstinacy	1	4	0.0001	0.0004	4.0114
praise		praise	15	86	0.0022	0.0061	2.8088
quarrel		quarrel	88	49	0.0012	0.0369	29.8332

END BOX

To cite this annex, use the Creative Commons method:
Olsen, Wendy (2021), *Systematic Mixed Methods Research for Social Scientists Annex of Practical Techniques*, Creative commons, URL https://github.com/WendyOlsen/SystematicMixedMethods

START BOX 8.2

How to Measure Keyness of Words and Then Examine the Discourses Reflected in the Words' Relative Prevalence in a Corpus

A keyness approach to deep linkage of quantitative and qualitative data could be based in part on Touri and Koteyko (2014). Their paper on corpus linguistics shows how to derive the frames from a text. The paper has been perceived mainly as guiding the analysis of journalism outputs, but it has much wider application. The paper presents framing whereas I see its main contribution as linking the 'keyness' and discourse concepts. A frame is a concept that enables us to break up the corpus of words into smaller parts (parse it) and then compare these words with those in a larger, definitive corpus for the same language overall. My own approach will be to work with individual words, rather than frames, because this simplifies our measuring-keyness step. We can move from the study of words or frames to the analysis of discourses in society. Many other possible methods can evolve out of such approaches.

I will introduce the Touri and Koteyko (2014) idea of keyness, and then go over the steps involved in operationalising it. My operationalisation is original, while the keyness idea is mainly attributable to Touri and Koteyko (2014: 605, 607, 2013).

Touri and Koteyko use software which compares the chosen body of text, called a 'corpus', with a broader corpus for that language. In their case, using English they chose to study a series of texts of speeches about the Greek Financial crisis of circa 2009-11. They used a version of frame analysis that focused upon key terms, because these key terms "tend to represent broader key concepts that become more prominent in the text." (Touri and Koteyko, 2014: 604). "Frames can be detected by searching for the presence or absence of certain keywords, which can function as an indicator of the conscious or unconscious judgement made by the communicator" (ibid.: 605).

This is a useful approach. The authors point out that if we carry out such an analysis deductively, there are risks of missing the deeper levels of meaning of the words. However when we are operating primarily from an inductive logic, moving from words to meanings is less problematic, they say. They use software packages to help summarise the key words and hence point out the frames in the texts.

In Touri and Koteyko (2014), keyness is measured through log likelihood, which relies on taking into account the number of instances of a word in the given corpus (compiled for a specific study) and comparing it to the number of instances in a reference corpus, which is representative of some norm (for example, the British National Corpus (BNC) comprising 100 million words used in spoken and written language) (Touri and Koteyko, 2014: 605)

It is easy to make the task of calculating the relative keyness of each term using a mixture of NVIVO and any statistical package. Just as Touri and Koteyko recommend, we are going to 'lemmatise' the words in our database. This means taking root terms, and then searching for all appearances of related words. We also will use concordances to develop coded segments of text surrounding each chosen word. Each word, in its turn, will get a 'keyness index'. This index will be high for words that are preponderant, low for rare words, and low for words which normally appear quite frequently but are not so frequent in this text. In other words, we develop a ratio of preponderances to achieve a relative frequency measure.

My own definition is closer to operationalising this concept with larger databases: the relative prevalence of words in one corpus of material over another. By providing an anchoring text I propose that keyness can be defined as shown below.

- Keyness: the relative prevalence of words in one corpus of material over another.

- Odds ratio index of keyness: the ratio of the odds of a word appearing in the studied corpus of texts versus in the overall language corpus of ordinary texts.

With these two definitions we can create a mathematical formula to help define keyness and show how it is derived. (I also avoided using logarithms.) The main formula is shown in Equation 1.

Suppose BBNC stands for the Baby British National Corpus, a large set of words derived from texts of ordinary English language in use (radio, conversations recorded in situ, newspaper texts, and so on). Let SSI stand for our Semi-Structured Interviews and the notation s for sample and n for overall sample size will be used in the equation.

The odds of a word appearing in a corpus are defined as the ratio of 'event' to 'nonevent', ie its occurrence count divided by all the other words' occurrences in the corpus. Thus in the SSI set, si/(S-si) is the ratio of a word's count to all other counts. In the overall BBNC, ni/(N-Ni) is the odds of that word appearing. We take the ratio.

$$\text{Keyness} = \frac{\frac{s_i}{S-s_i}}{\frac{n_i}{N-n_i}} \quad \text{(Eq. 1)}$$

For each word numbered 'i' in the set of all words in the BBNC corpus. Some words in the corpus do not appear in your dataset, so they drop out. Some words are unique to the new dataset, such as place names and pseudonyms, so keep these as if they were high in keyness.

In summary, the proposed keyness analysis offers a form of corpus linguistics that can be useful to a wide range of disciplines beyond, and yet including, linguistics. Keyness analysis offers a dispassionate way to use all the semi structured interview data or the focus group data from a project. It can be used for any large dataset involving sentences and other utterances. When I use it for one interview, I obtain just a short list of high keyness words. When I use keyness analysis on 39 interviews, each of about 30-40 minutes in length, I get just 233 key words based on a corpus of 47,000 words.

I show how to do it in three steps.

Step 1. Conceptualise your Keyness Measurement and the Interpretation

According to Touri and Koteyko (2014), keyness is the extent to which a word or phrase is central to the topic of an interview or segment, above and beyond its normal usages. Touri leaves out grammatical inking words like 'and' as well as 'lemmatised' relatives. For example (see Appendix 1), taking cultivation, sharecropping and kuulie, we have as lemmatised relatives cultivator, cultivates for the first; share-cropped and sharecropper for the second, and kuulie work and kuulies for the third. The English grammatical words we drop are 'in' 'he' 'we' 'are' and 'just' and some others. The key pronouns will reappear in a later stage when we add concordances (Touri and Koteyko, 2014) . The 'keyness' of the phrase depends not just on its own prevalence among the interviews. Instead we standardise this, dividing each prevalence by the overall prevalence found in a basic corpus (body, or encyclopedia) of materials for that entire language, in this case English. The Corpus we use for English is the BBC Standard English one, and for each keyword we seek its absolute prevalence in that Corpus. Dividing these by the sum of all the relevant prevalences, we have its standardised prevalence. Then we divide the Interview Keywords' Local Prevalence by BBC English Prevalence, giving a ratio term. (This is not too different from the log-odds or odds ratio measure of the probability of this term appearing in this set of interviews, given the size of the whole set of texts.) At Appendix 3 we show the result for a small sample extract.

After choosing our 'keywords' with high keyness, using NVIVO we do a 'concordance' of them, in which we look at the words appearing with them. Thus for the key word 'studies' we will have 'he' associated with studying but we may not find occurrences of 'she studies'. Based on my small pilot research with keyness, I think 3 sentences is a good size of concordance for the kind of rural respondents we had in India, Bangladesh and Sri Lanka. In a different context, you might choose a larger segment or a smaller segment for concordance. Without 3 sentences, we will not get much context nor will we understand the import and meaning, or connotations, of the key word.

In Touri and Koteyko's terminology, we look for 'frames' when we seek to get the phrases or sentences coded as a larger concordance. A frame is more than a single word. A frame contains a unit of meaning. EMPIRICAL FRAMES FOR MARRIAGE, for example, would be the list of coded segments we obtain after carrying out an automated annotation of 3-sentence segments about terms linked to the marriage topic.

Next we group the most key words together. Remember, **'key' means having a relatively high prevalence, while 'common' would mean merely having an absolutely high prevalence.** We group each term one once or twice as required, through a manual interpretive procedure, to generate a gathered set of named discrete discourses in which the respondents were engaging. Discourses overlap and we are multiple-coding many segments of the original text.

Each quote can now be interpreted in terms of whether it is consistent with the usual (dominant) usage of that discourse, or in some cases whether due to intertextuality the phrase was used in a novel, sarcastic, questioning or competing way. This analysis is done selectively due to the possibly overwhelming amount of work involved.

Step 2. A Summary Plan of the Coding in Preparing a Coded Corpus for Interpretation

First, compare your data corpus with the wider definitive corpus for that language.

The computerised mechanisms in qualitative coding software and in a spreadsheet are very straightforward. Now there is another step where more human thought and careful intervention is required.

We intervene in the coding process, choosing which words have high enough keyness to matter, so that the coding does not exceed some limit of feasibility, here set at 800 key words. The key human decision is which cutoff level to use.

The Stages of Human and Computer Mediated Coding

The scope of the project will not necessarily go beyond a feasible scope if the keyness cutoff is chosen to suit the current means of the project team.

Here is an applied example. If you have 500 keywords (that is, words of length 5+ according to the decision I made) in one interview, then in twenty interviews that have a lot of topic overlap, there may be 1500 distinct keywords in all. Among these, pick the top half, or 750. These have all got a reasonable prevalence, and we choose as the cutoff the keyness level at the 750th ranked word. Now the coding of their 3-sentence surrounding material segments will be intense, and often carried out by hand. Two keywords may overlap in the final coding coverage. Obvious examples could be love/marriage (in South India), work/pay, and mother/duty.

You may use the NVIVO software (see https://www.qsrinternational.com/, accessed 2020). To initially code the concordances of each of the 750 key words use NVIVO Query > Create. Copy the key term in from Excel. Spread the query to 3 sentences. Click 'save as a node'. Name the query. Ensure it will hit all interviews, not just the one we've coded so far. Save the Query. Run the Query. Do this 750 times. You now have 750 new nodes. (If you use other software you may automate this process.)

You can now search these nodes for the material overlapping with a particular discourse by creating a case-node for each discourse, then using a new Matrix Query.

Step 3. Link the Coding of Words and Frames to the Discourses (Illustrated)

Sample of Translated South Indian (Telugu) Discourses and How Accessed Via Key Words

Key Word	Discourse, meaning, discussion
Cheat	

Property Thousands …and right, lady, connection	1. In the **discourse of property inheritance**, the rights of recipients and the debate over cheating makes reference to law as well as family connections. 2. A **discourse of family obligations** states that family members should be loyal to their parents and lineage. Resistance, desertion, and exceptions are discussed at length. These discussions act partly to discipline those now living in families.
tenant agricultural	3 In the **discourse of working** on the land for arable farming, the hard work is mentioned with pride. A quote illustrating that other forms of work are also admired, a man speaks admiringly of a good woman doing "clay work, open well work and tank work"
Struggle Connection Right	1. *(sic)* In the **discourse of family obligations**, inheritance exceptions are hotly debated, and some go into court, and last for decades. Both families support their struggling parties who champion their rights, obtained through connection and family lineage, although they are also pitted against other claims of family rights.
Husband lady	4. **Gender roles** are often referred to via individual people's family roles. But when a person is speaking with disapproval about a woman, the term 'lady' is used (in translation) without a kinship term attached and this lack of friendly/familiarity sends the message of disapproval. Again this acts as a disciplining

	message. We can denote these usages of both disapproval (here) and fictive kin suffixes –amma –akka etc. as approval, as the **normative familial gender roles discourse.**

Note: this example is from my 2006-7 study of land tenancy and work and gender in South India. The examples here are from a single interview.

I will now summarise the method proposed. To use a keyness approach to a mixed-methods project, you might follow the steps shown below.

Flowchart of the Coding Tasks for Deep Linkage

1 NVIVO & EXCEL
- check keyness of all main terms
- code the key terms with their concordances

2 Compare with Statistics
- check how the keyness statistics can be done, if at all.
- A. if you seek explanations, examine what discourses say, and code patterns you observe.
  - For example you might do Matrix Query to compare men and women's use of a particular discourse.
- B. if you seek descriptions, compare the verbs, metaphors, entities and features as named in competing discourses
  - For example you may create tables or bar charts of the prevalence or modal frequency of each key word, by social groups.

3 Notice the Discourses
- divide up the key words into topics, and examine the data; decide what dominant and marginal discourses exist
- choose only those that you have time to focus puon
- code up these main discourses manually
- notice which are used with high/low salience (that is, were speakers aware of the discourse?)

Pick out Unique Cases
- For each interview, notice the intentions of the speaker.
- Are they traditional/perpetuating? or egalitarian/innovative?
- Is there intertextuality? if so, how do agents innovate?

Bibliography of Works Used in Creating This Online Annex:

Eliason, Scott R., & Robin Stryker (2009) Goodness-of-Fit Tests and Descriptive Measures in Fuzzy-Set Analysis, *Sociological Methods & Research* 38:102-146.

Heale, R., and D. Forbes (2013), "Research Made Simple: Understanding triangulation in research", *Evidence-Based Nursing* 16:4 98 doi:10.1136/eb-2013-101494.

Morgan, J., and W.K. Olsen (2011) "Aspiration Problems for the Indian Rural Poor: Research on Self-Help Groups and Micro-Finance", *Capital and Class*, June, 35:2, 189-212, doi: 10.1177/0309816811402646.

O'Cathain, A., E. Murphy, and J. Nicholl (2010) "Three Techniques for Integrating Data in Mixed Methods Studies", *British Medical Journal*,341: , doi: 10.1136/bmj.c4587.

Olsen, W.K., and J. Morgan (2010) "Institutional Change From Within the Informal Sector in Indian Rural Labour Relations", *International Review of Sociology*, 20:3, 535-553, on "The Regulation of Indian Informal Capitalism", DOI:10.1080/03906701.2010.51190.

Olsen, Wendy, and John McLoughlin (2016, 2019), Fuzzy Set Goodness of Fit Tests Version 1 or Version 2 [*Software package*], URL https://github.com/WendyOlsen/fsgof, accessed 2019 (Version 1) and for version 2, 2019, see URL https://github.com/luminous-badger/cdnec, accessed 2019.

Ragin, Charles, and Sean Davey (2014) *fs/QCA [Computer Programme]*, Version [3.0]. Irvine, CA: University of California. The instructions for loading are at: http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml accessed 2020.

Touri, M., and N. Koteyko (2014) "Using Corpus Linguistic Software in the Extraction of News Frames: Towards a dynamic process of frame analysis in journalistic texts", *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2014.929878.

Acknowledgement: I am grateful to John McLoughlin who helped with implementing the keyness formula in the 'Baby' British National Corpus.