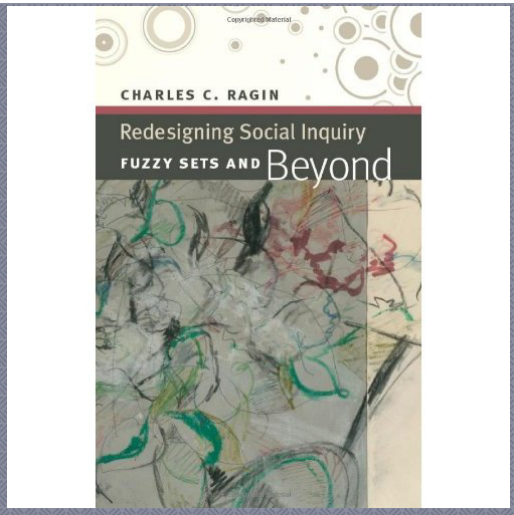
What Is …
Fuzzy Set Analysis?

A MODE OF ANALYSING DATA USING SET MEMBERSHIP AND CONFIGURATIONAL ANALYSIS

FEBRUARY 2012
METHODS AT MANCHESTER

WENDY OLSEN

Ragin distinguishes necessary cause from sufficient cause in fuzzy set analysis; whereas some statisticians say we cannot know that anything is causal.



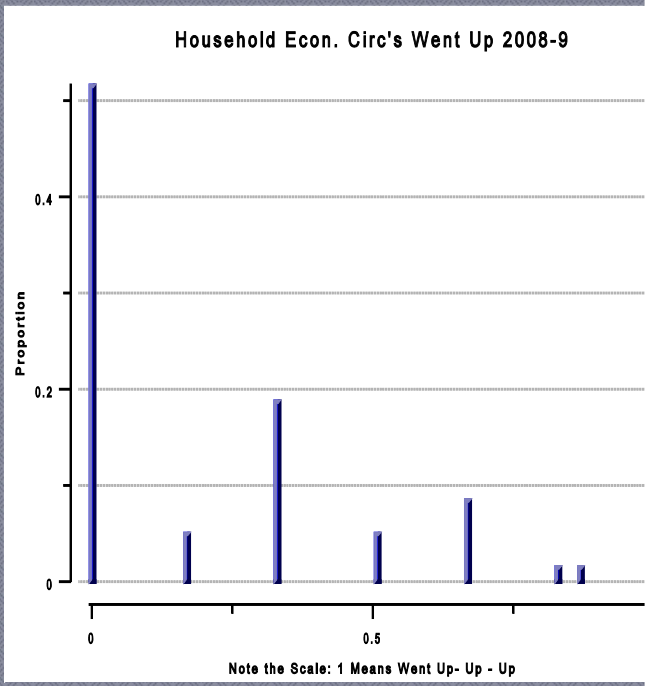
Plan of this Talk

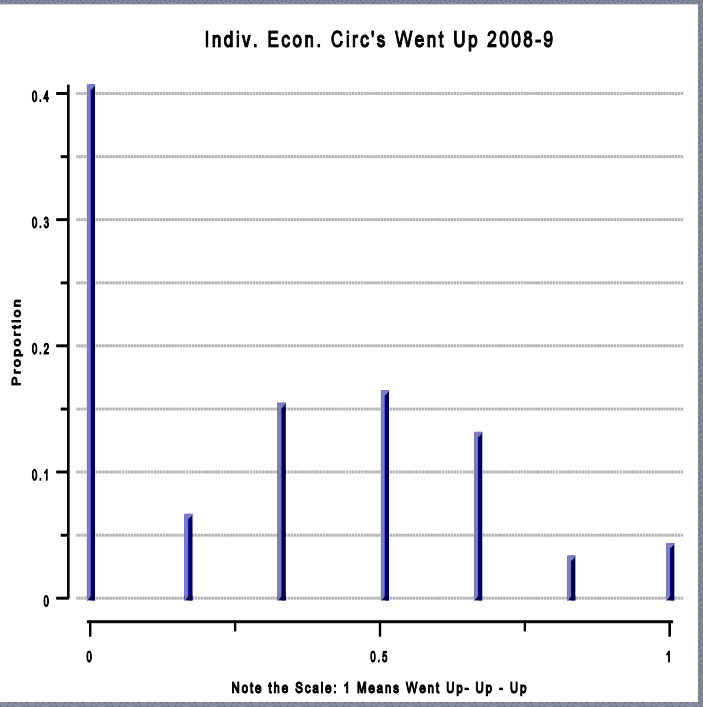
- I hope to introduce 1 Fuzzy sets as a measurement method
- 2 3 reasons to use Fuzzy Sets in qualitative research
- 3 The mode of logic used in fsQCA
  - 3.1 the subset relation
  - 3.2 S-consistency (vector X sufficient for Y)
  - 3.3 N-consistency (X necessary for Y)
- 4 conclusions and exemplars

1 Fuzzy sets as a measurement method

- Fuzzy set membership ranges from 0 to 1 in sets defined conceptually or qualitatively. E.g. 'Living' / 'Dead'
  - Fuzzy set version could be:
    - Living / suffering / damaged / incurable / palliative care only / dead
    - 0 0.2 0.3 0.7 0.9 1.0
    - This could be used in health research, as the outcome that matters is being anywhere in the set of Dead.
    - Then one does not have to await post-mortem to get definitive results on the obstructive effect of the treatment, which inhibits death effect of disease.
  - We calibrate the set using a 0.50 level either as a value on the spectrum of ordinal levels, or as a value to be avoided. 0.50 is both 'in' and 'out' of the set of dead patients.

Fuzzy Set Measurement





Households' and Individuals' Economic Circumstances Went Down (Shown at the Left as the Predominant Outcome)

Another Application

- Fuzzy Set Theory Movement in the Social Science, W.A. Treadwell, Public Administration Review, 55:1 (***1995***), 91-98
- **"a fuzzy logic conclusion is not stated as either true or false, but as being possibly true to a certain degree. The degree of certainty is called the "truth value." Fuzzy set theory uses only the numeric interval of 0 to 1:**
  - **FALSE: TruthValue= 0**
  - **TRUE: Truth Value = 1**
  - **UNCERTAIN: 0 <Truth Value < 1**

Peer Fiss's work
Cooper's work (see also Blackman at Durham)

- Fiss argues that organisational strategy and outcomes are best placed on fuzzy sets. Also that there are **core** causal groupings which are really central to good organisational strategy outcomes. Other causes are merely peripheral. He makes this an empirical question.

- Cooper studies mobility, class, education. He argues that fuzzy sets offer a better mode of argument than continuous measurement for each variate.
  - Byrne argues for 'variates' not 'variables' in contexts where we have Small-N.

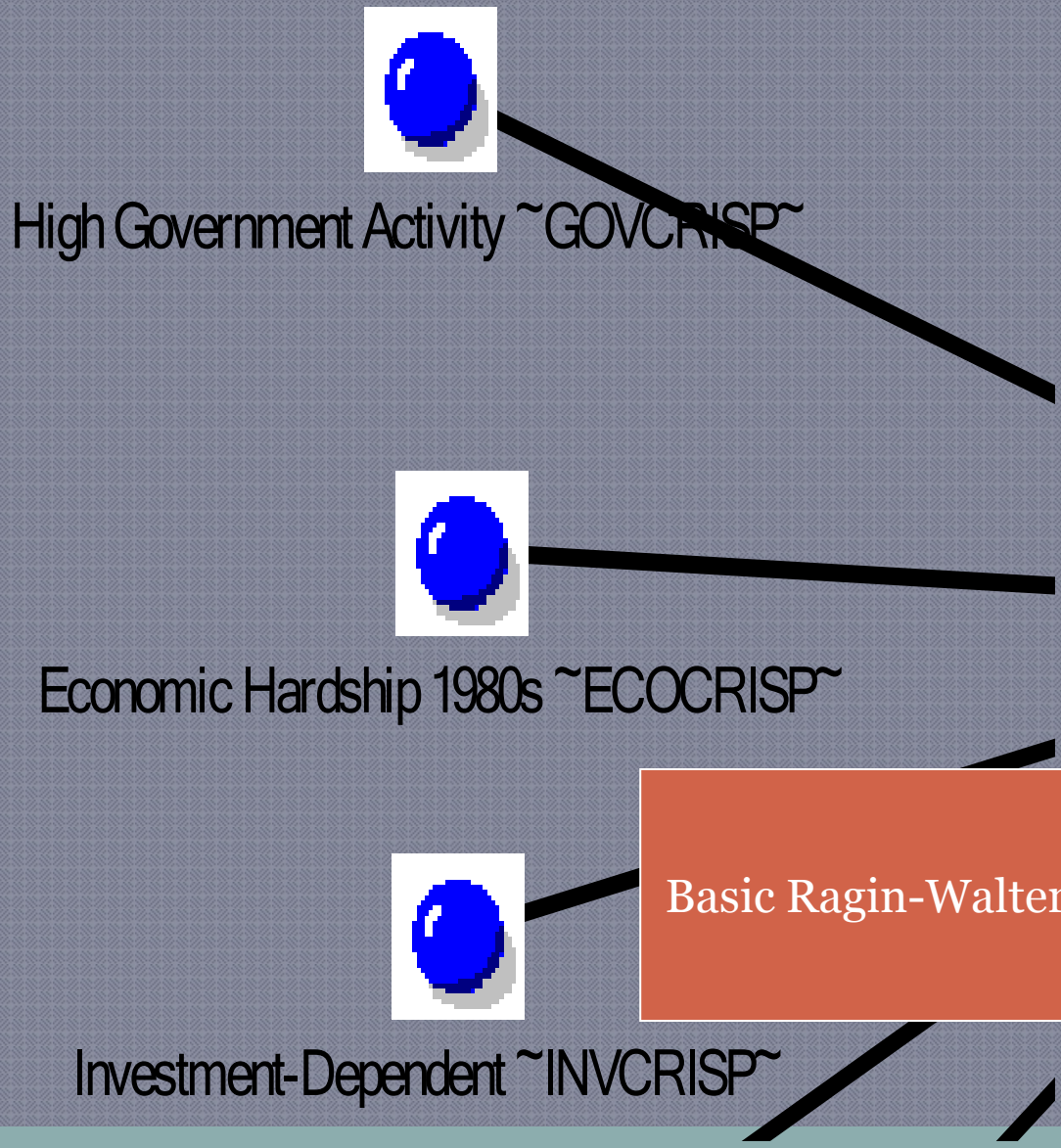
Management applications

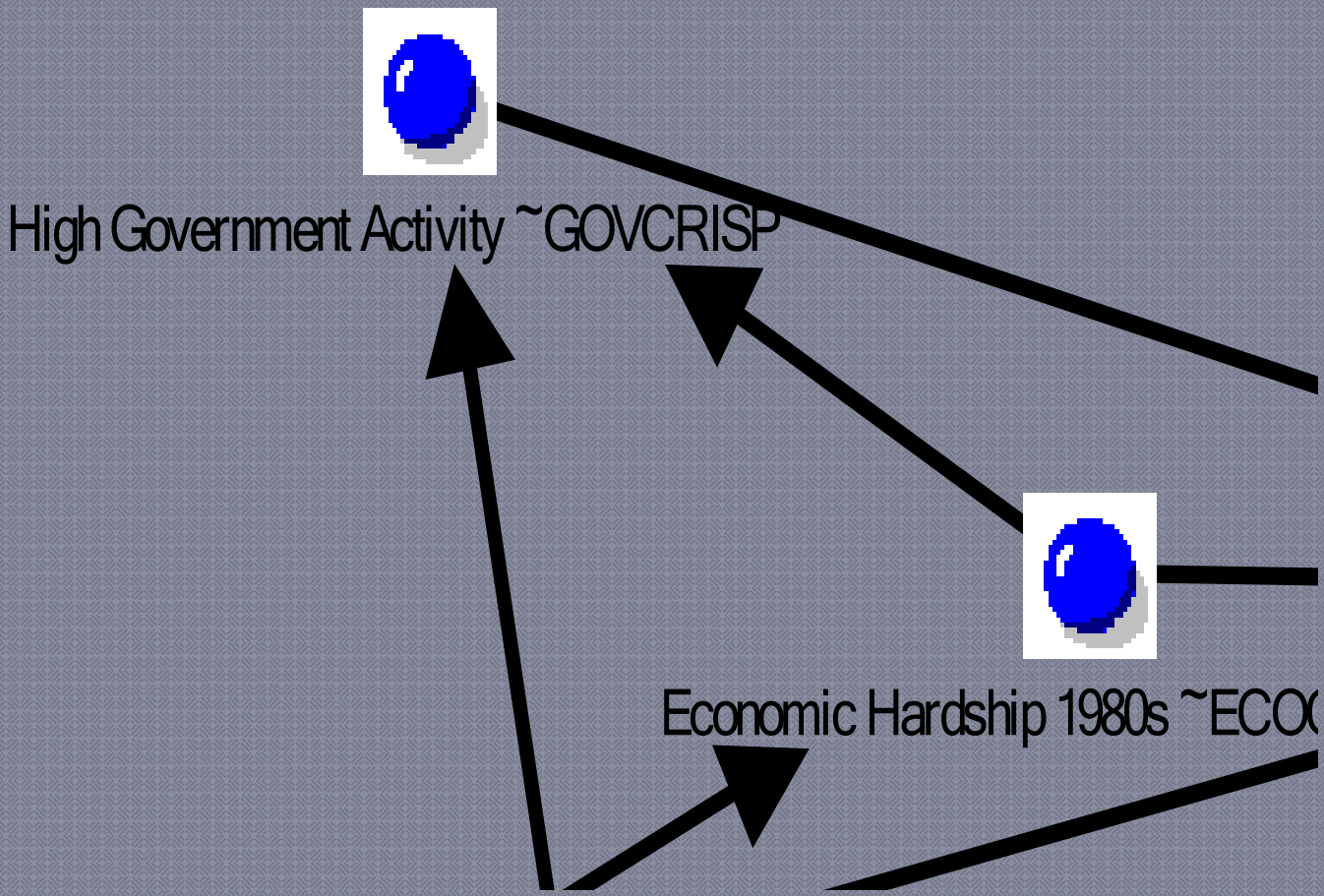
Good Outcomes, e.g.
Y= Fuzzy Set of Growth of the Firm

Values Human resources approach Legitimacy of the system	Mode of leadership Strategy (requires indicators) Partnerships	Resources such as people, skills, inputs of training	→ Y

Ragin's Own Applications (FSSS 2000)

- Country level measures of protests against the IMF Structural Adjustment Programmes
- IQs
- Breakdown of democracy in the inter-war period in Europe and nearby
  - E.g. Italy, Germany vs Belgium?, France?, UK
  - He made a fuzzy set of the breakdown of democracy
  - He also has causal factors, also fuzzified.


High Government Activity ~GOVCRISP~
Economic Hardship 1980s ~ECOCRISP~
Investment-Dependent ~INVCRISP~
Basic Ragin-Walters Model

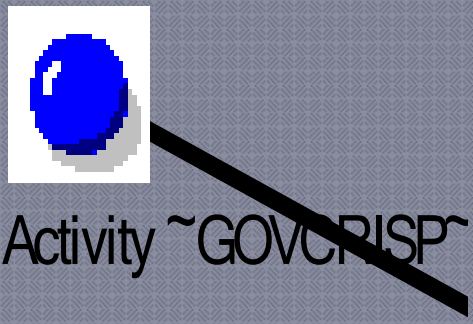


Modified Ragin-Walters Model

A model of growth

ECONOMIC OUTCOMES WITH STRUCTURAL FACTORS STRONGLY DETERMINING THEM PERFORM VERY WELL IN FUZZY SET ANALYSIS

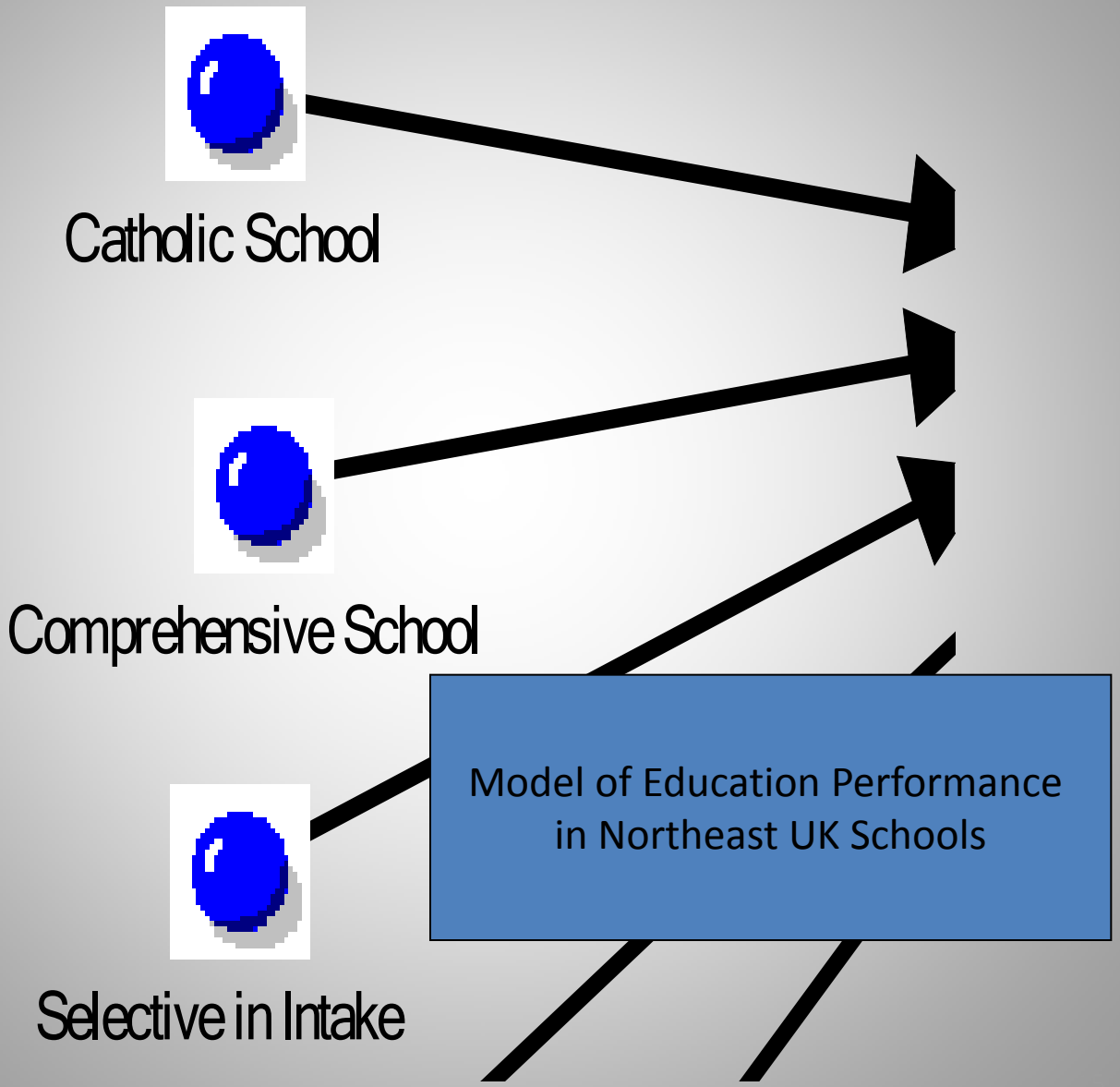
THE USUAL PATHWAYS OPERATE:
OVERALL GOVERNMENT SPENDING;
FOREIGN DIRECT INVESTMENT;
LIBERALISATION



Simplified Post-1990 Growth Model

We tested school A-level and GCSE results by social deprivation, Selective, etc.

- See Byrne, chapter in Byrne & Ragin, Handbook of Case-Centred Research, London: Sage, 2009.
- See also Byrne et al. article in *Theory, Culture and Society* on causality.
- See my paper in *Handbook*, 2009.


Catholic School
Comprehensive School
Selective in Intake
Model of Education Performance in Northeast UK Schools

Measurement in Sage volume

- In Set theory we call the fuzzy set a measure of membership in the set.
- It is then thought of as a cardinal measure. I.e. On a continuous scale.
  - See Smithson & Verkuilen, 2006
  - Or Longest and Vaisey, STATA FUZZY 2009
- Another way to construct a fuzzy set is to develop it from a continuous measure, such as GDP per capita. Decide on 'poor' / 'Non-poor' as the set you are working with. Choose a 0.5 point, and calibrate the set.
  - Transform values onto 0-1 scale with a particular GDPpc = $700pcpa as cutoff point.
  - This calibrated set is a monotonic transformation of the original GDPpc measure.

Key references for technically skilled folks

- Smithson, M. and J. Verkuilen (2006). *Fuzzy Set Theory: Applications in the social sciences*. London, Sage.
- Mahoney, J. (2008) Toward a unified theory of causality. Comparative Political Studies. 41, (4-5), 412-436
- Longest and Vaisey, article in STATA Journal, 2008. They created the FUZZY command which does QCA and fuzzy set calibration in STATA.
- fsQCA freeware, see http://www.u.arizona.edu/~cragin/fsQCA/software.shtml
- Download version 2.0 and the Manual

Basic Rules of Fuzzy Sets

Appendix A
Fuzzy Set Properties

The algebraic properties—commutativity, associativity, distributivity and complimentation—of fuzzy sets are the same as for ordinary sets (i.e., binary sets with the two elements "0" and "1") using Boolean logic. The rules for intersection and union and compliment are the same. The primary difference in properties between ordinary sets logic and fuzzy sets logic are the properties of noncontradiction and exclusion.

- Commutativity: $A \cap B = B \cap A$
  $A \cup B = B \cup A$
- Associativity: $(A \cap B) \cap C = A \cap (B \cap C)$
  $(A \cup B) \cup C = A \cup (B \cup C)$
- Distributivity: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- The universe of elements in fuzzy set theory is any number in a closed interval from 0 to 1.
- A fuzzy set has a membership function with not only values of 0 (does not belong to) or 1 (belongs to), but any number in the interval 0 and 1 (For example, 0.3, 0.651, 0.98...that represent grades of membership.)
- The following two fuzzy sets (A,B) are used to illustrate.

$A = \{.7 \quad .4 \quad 0 \quad .5 \quad .2 \quad 1\}$

$B = \{.3 \quad 1 \quad .4 \quad .9 \quad 0 \quad 1\}$

Notice: Cases are assumed to exist, to be comparable, not to be nested...

Combining Fuzzy Sets

- The following two fuzzy sets (A,B) are used to illustrate.

$$A = \{.7 \ .4 \ 0 \ .5 \ .2 \ 1\}$$

$$B = \{.3 \ 1 \ .4 \ .9 \ 0 \ 1\}$$

- Intersection of A and B uses the minimum rule of selecting the smaller of the two elements

$$A \cap B = \{.3 \ .4 \ 0 \ .5 \ 0 \ 1\}$$

- Union of A and B uses the maximum rule of selecting the larger of the two elements

$$A \cup B = \{.7 \ 1 \ .4 \ .9 \ .2 \ 1\}$$

- The compliment of A (not A) is 1 minus the member in A:

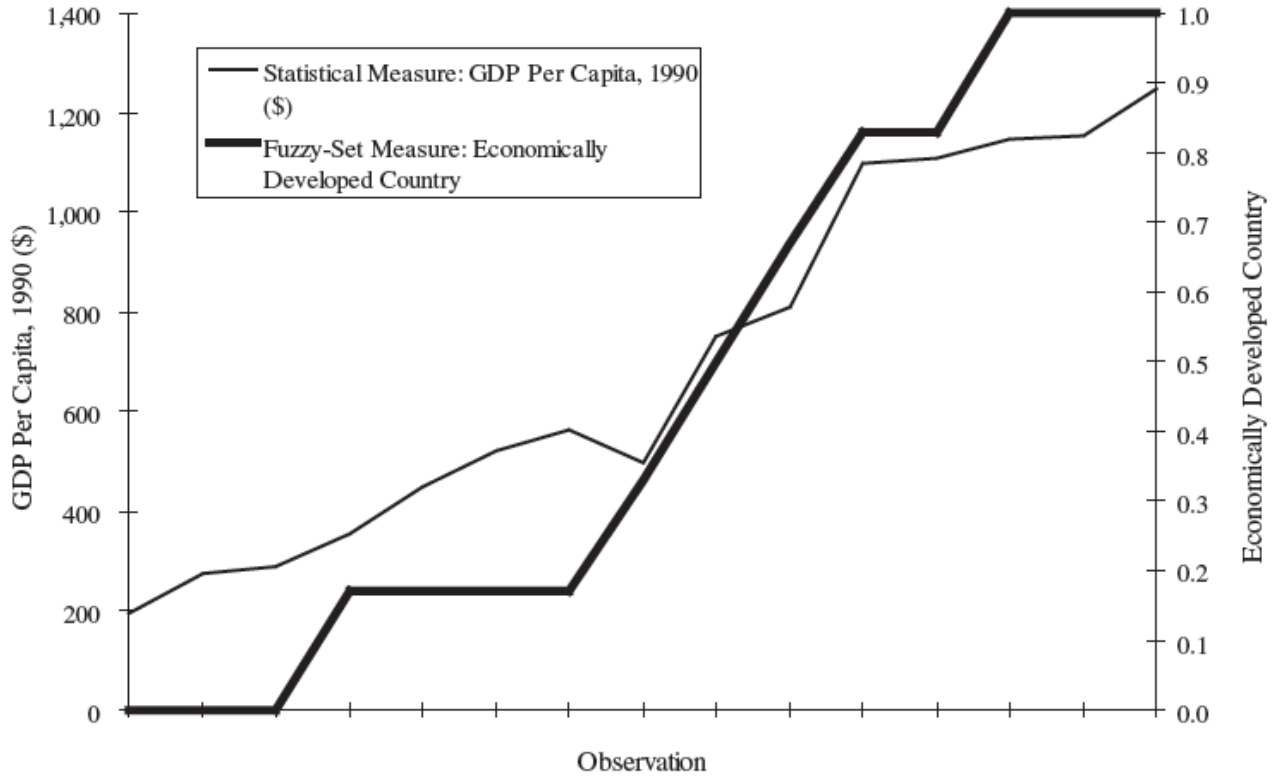
$$A^c = \{.3 \ .6 \ 1 \ .5 \ .8 \ 0\}$$

Applied Uses of The Sum of Sets

- The human development index and other measures of human capabilities
  - Martinetti (HDI)
  - Anand (BHPS)
  - Neff (subjective well-being)

  - Please note that there is no need to be massively universalistic or grossly macro about this. You can measure the 'Multiple Deprivation Index' this way too – it will not still be decomposable, but it will respond to specific areas of deprivation differently-
  - Depends on whether you use the UNION (SUM, OR) or the INTERSECTION (AND) function to link measures.

Another example of fuzzy measurement



The resulting raw truth table…

TABLE 2: Fuzzy-Set Scores for Causal Factors

	Dense Indigenous Population	*Labor-Intensive Estates*	*Mineral/Tropical Exports*	*Strong Liberals*	*Strong Conservatives*
Argentina	0.17	0.17	0.33	1.00	0.50
Bolivia	1.00	0.83	1.00	0.17	0.67
Chile	0.33	0.83	1.00	0.83	0.50
Colombia	0.67	0.50	1.00	0.83	0.83
Costa Rica	0.00	0.00	0.83	0.67	0.00
Ecuador	1.00	1.00	0.50	0.33	0.67
El Salvador	0.83	0.67	1.00	0.83	0.50
Guatemala	1.00	0.83	1.00	0.67	1.00
Honduras	0.67	0.17	0.50	0.33	0.17
Mexico	1.00	1.00	1.00	1.00	1.00
Nicaragua	0.83	0.33	0.33	0.67	0.67
Paraguay	0.50	0.50	1.00	0.33	0.67
Peru	1.00	1.00	1.00	0.33	1.00
Uruguay	0.00	0.00	0.17	1.00	0.00
Venezuela	0.50	0.33	0.33	1.00	0.33

# These authors also used statistical methods to illustrate the result:

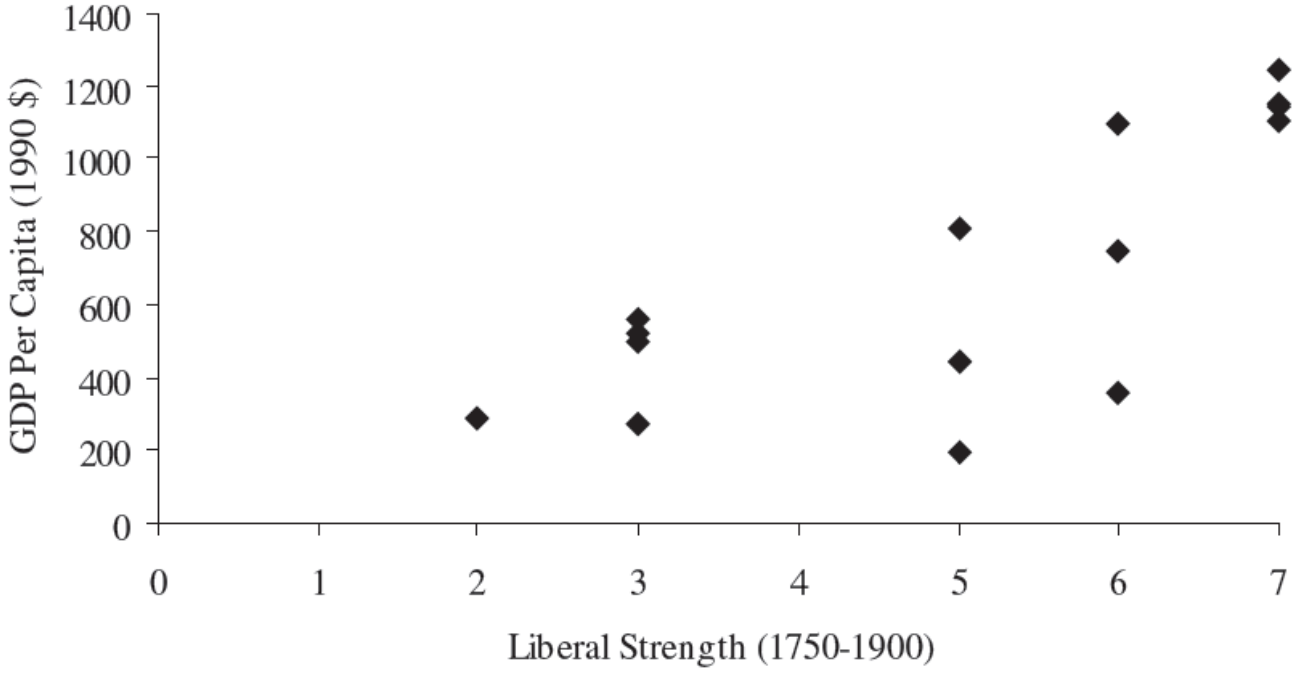


Figure 3: Liberal Strength (1750-1900) Versus Gross Domestic Product (GDP) Per Capita (1990 $)

# These authors also used statistical methods to illustrate the result:

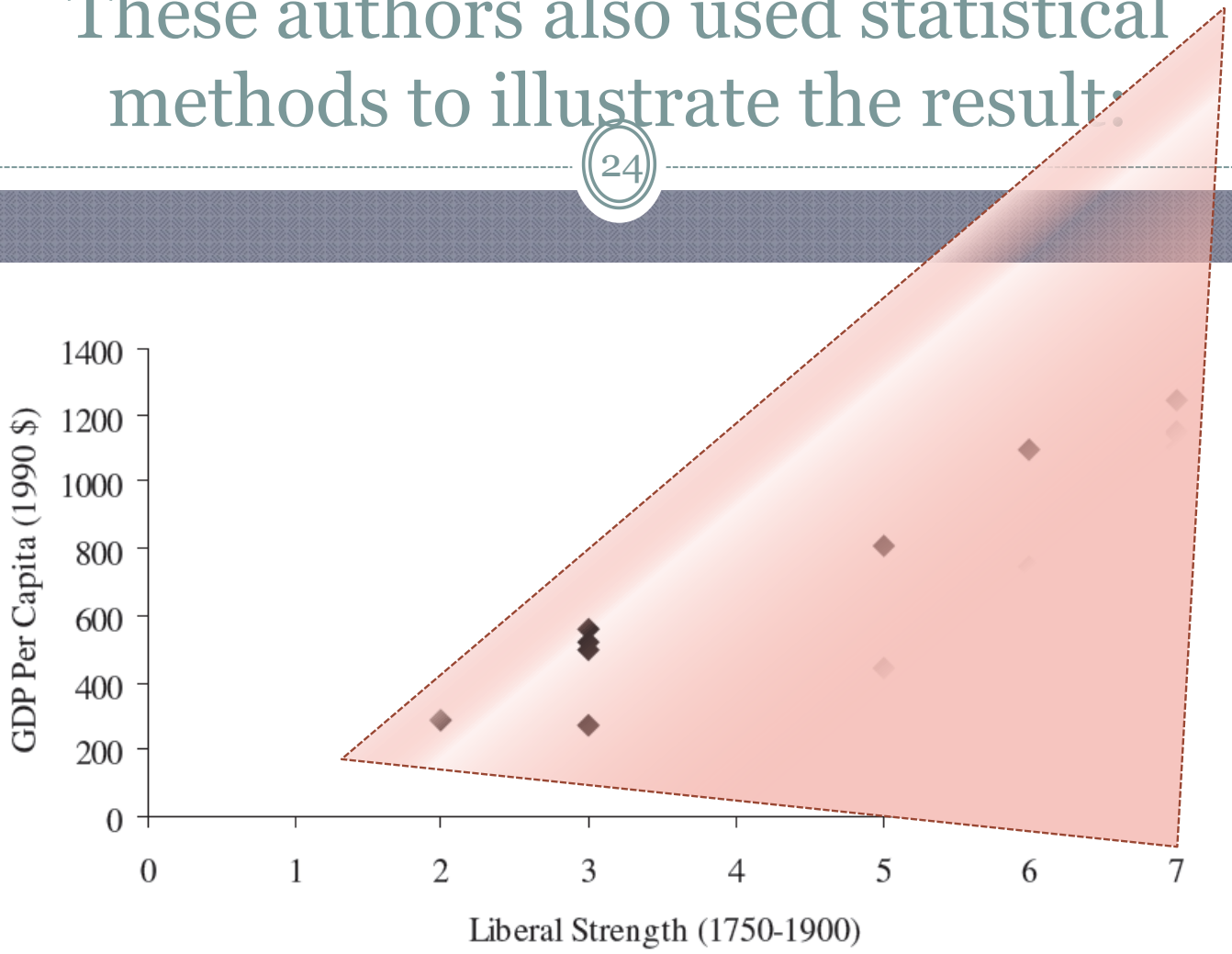


Figure 3: Liberal Strength (1750-1900) Versus Gross Domestic Product (GDP) Per Capita (1990 $)

This shape tends to illustrate necessity

- The example has moved from the fuzzy set scale to the actual GDP pc scale to be less controversial.
- They used fuzzy set QCA to explore and examine the data.
- They found out which configurations mattered for their main outcome, which is being high-income today at country level.
  - In the end one factor dominated
  - But in the study, five factors were considered.

Section 2: Three Reasons to Use Fuzzy Sets in Qualitative Research

- 2.1 Firstly when we have small N and there is not a random sample, it is not sensible or realistic to do statistical inference.
- 2.2 Secondly when we have a whole population, or a part of a population of cases, we need something other than statistical inference.
- 2.3 Thirdly we can use Expert Interviews or Stakeholder Analysis to develop fuzzy sets as a measurement method.

Hypotheses Well Suited to Fuzzy Set Analysis

- The supply of clean water in pipes is restricted among poor households in its ability to provide access to clean water because they can't afford to buy the service.
- The supply of clean water in pipes is not taken up much even by middle income groups due to intermittent supply, i.e. Some is as bad as none.

Why does QCA play a role?

- QCA allows us to compare the success in water-management of different social configurations. Rural/ periurban, high- or low-income, piped or well, with charging/free, via private company, or via government... Etc etc.
- We are not looking at Does X work for Y?
- We are looking at ‘in what circumstances does X seem to be able to produce Y’

Some QCA Concepts in this Scenario

- necessary cause;

- sufficient causal mechanism

  - 3.1 the subset relation
  - 3.2 S-consistency (vector X sufficient for Y)
  - 3.3 N-consistency (X necessary for Y)

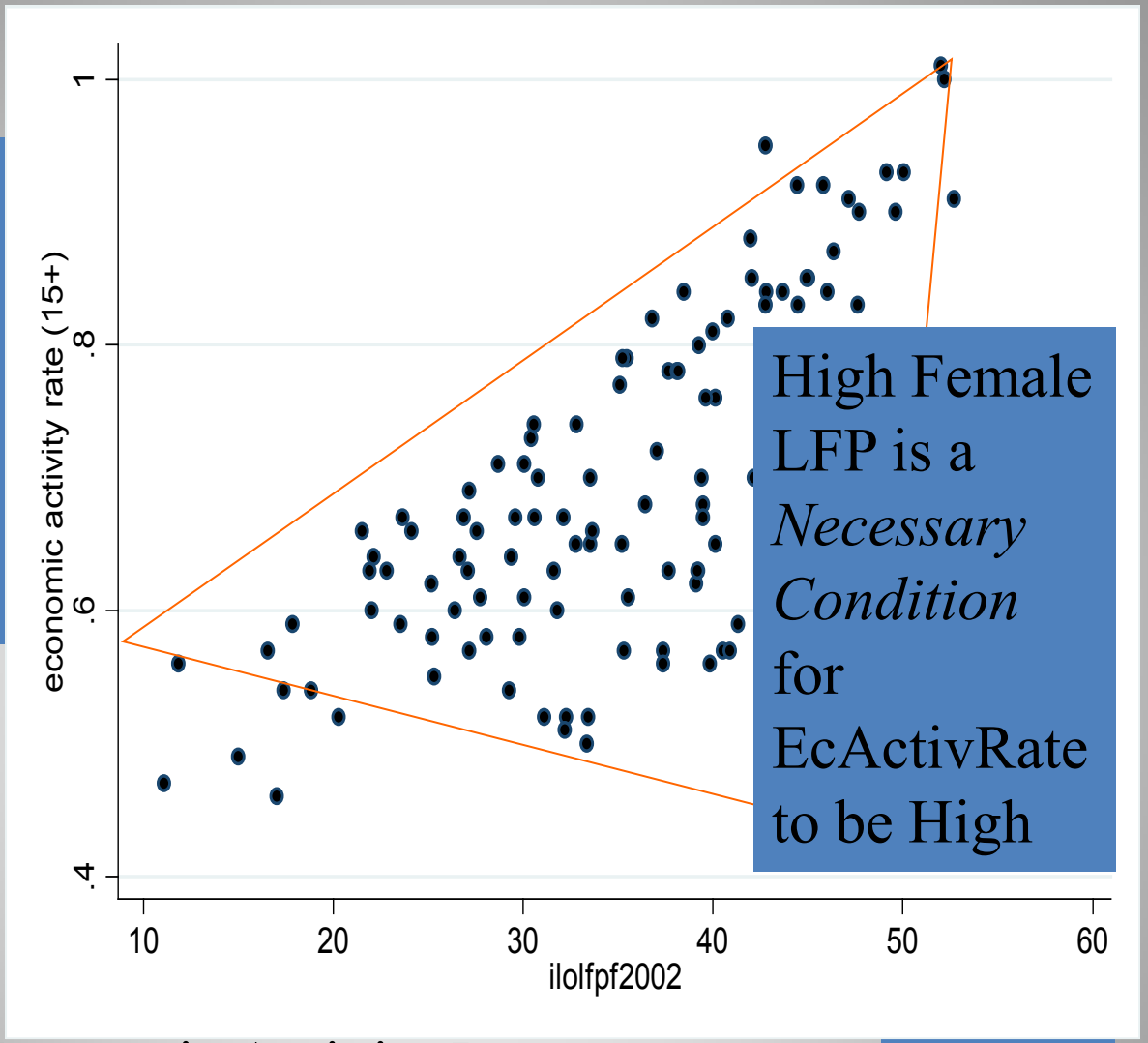
The formula for necessary cause

- It is claimed that if X is a subset of Y, in fuzzy set terms, then N-consistency is a measure of how far X is necessary for Y.
- N-Consistency is an adjusted measure of the sum, for all cases, of the union of X with Y. The adjustment is the sum of all scores of Y.

Necessity appears in social situations

- For example, among a variety of capitalist countries, the labour force participation rate depends upon the women's labour force participation rate.

Econ. Act. Rate = LF/Pop.



High Female LFP is a *Necessary Condition* for EcActivRate to be High

Rate

The Economic Activity Rate Age 15+ is constrained by the Women's Labour Force Participation Rate 15+ (data for

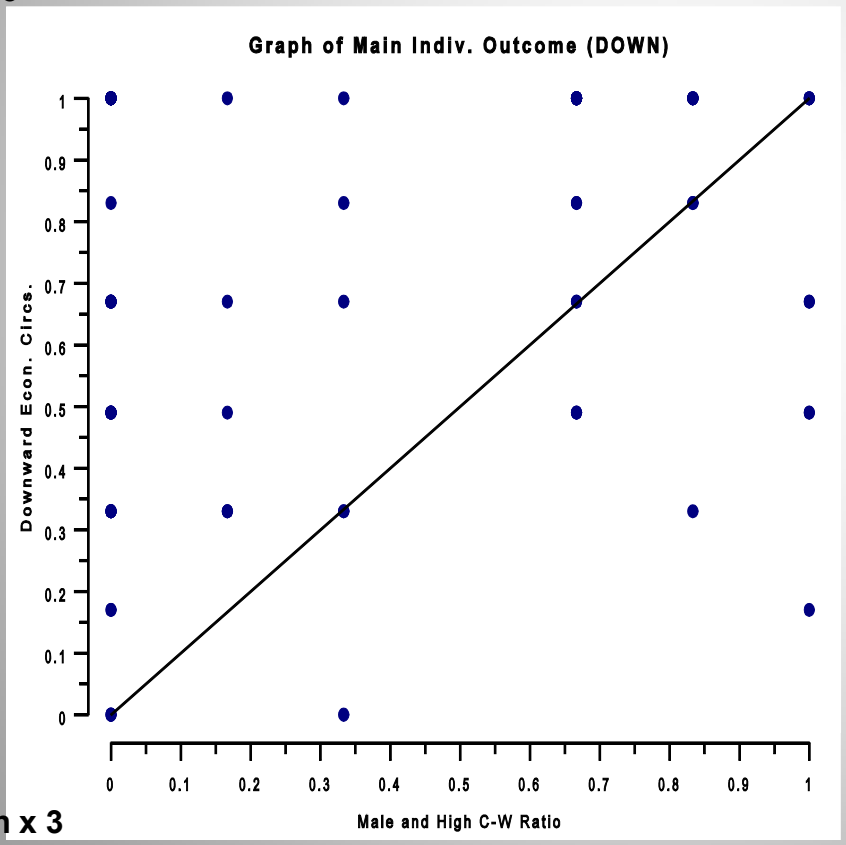
Testing for Sufficiency

- A cutoff of 0.8 is advised by Rihoux and Ragin (2008) and Ragin (2009).
- ***Notice that in micro data, this cutoff may not be reached, as shown here:***
  - ***Horizontal axis is a vector composed of two fuzzy sets at household level (male respondent and a high consumer-worker ratio, i.e. A lot of kids in the household)***
    - ***Male respondent is a binary condition. Binaries can easily be combined with fuzzy sets using Boolean algebra.***
  - ***Vertical axis is a fuzzy set of how they experienced the shocks of 2008 in Indian city of Chennai.***

- ***The failure of a test for sufficiency still leaves us interesting avenues to explore. We can add more elements to the configuration in case something else is affecting the likelihood of being in set Y.***

The QCA Method: Sufficient Cause

Up x 3



Down x 3

Strong dominance of upper left triangle suggests that X is sufficient for Y in most cases. (N=91)

Next test for sufficiency of Cause of W for each CONFIGURATION or VECTOR X

- The vectors differ in complexity e.g.
  - A vs a,
  - Ar and AR vs. R
  - ARF vs. Arf , arf, aRF, etc.
  - ARFL vs. Other permutations with 4 elements
    - I call these vectors but they are also called configurations.
    - That is because for each one, the 'yes' group ARFL has N1 cases, ARFl has more cases, and so on, each group with Y= yes or Y = no, or Y high/low fuzzy set membership.

- As a result there are many, many permutations. Software helps ease the decision problem.
- fsQCA freeware picks out the relevant configurations sufficient for High Y.

Illustration and Challenges

- Use Lam and Ostrom to Illustrate. Convert their table to numbers. In fsQCA set the S-consistency cutoff at 0.80 and frequency >= 1.

Data – Raw – Nepal Irrigation Success in W

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

Number of cases

The number of **rows** here is 11. This is the number of configurations, allowing for some contradictory configurations, i.e. counting as a single configuration both the aRfLC combination with W and NOT-W outcome. That is, aRfLCW and aRfLCw.

fsQCA Results

- File: C:/Users/Wendy/Documents/New Folder (2)/waterAlternPaper/LamTableW.csv
- Model: w = f(a, r, f, l, c)
- See next page for results.
- This is also what Lam & Ostrom report.

Results, Crisp Set Quine Algorithm

- Notice: here ~ means ‘not’.

```
                raw          unique          S-
                coverage     coverage        consiste
                ----------   ----------      -----
a*r*c           0.50         0.250           1.00
~a*~f*~l*c      0.17         0.167           1.00
a*r*f*~l        0.17         0.083           1.00
r*f*l*c         0.33         0.167           1.00
solution coverage: 0.917
solution S-consistency: 1.000
```

Conclusions

- Fuzzy set membership functions mean different things depending upon the discursive context.
- Their use in configurational analysis and among realists tends to be to interpret pathways of causality.
- The empiricist problem with this is that sufficiency (measured by consistency) is only 1.0 if the coverage is <1.00.
- Therefore we have to have a non-deterministic approach to causality
  - Generative approach – realist – very plausible!
  - Complexity-based approach – very sensible!
- This makes error not longer error, but information.
- The necessary cause shape differs from the sufficient cause shape.
- Many, many applications exist (300+ peer reviewed journal article applications so far)

Get Involved

- Three ways
  - Working papers and peer reviewed journal articles:
    - www.compasss.org
  - Join an email list to ask/see queries and answers:
    - www.jiscmail.ac.uk >> QUAL-COMPARE
  - Join a training course
    - Shortage of training

Followup Reading

- A paper covering QCA from a realist viewpoint, in an applied setting, presented a Workshop in Bonn in 2012 by Wendy Olsen
  - http://dl.dropbox.com/u/13339002/conceptNote2011OLSEN.doc
-
- The Powerpoint slides from that lecture in London cover the realist basis for configurational QCA more:
  - http://dl.dropbox.com/u/13339002/WaterPresentationOlsenLondonOct2011.ppt

References I

- Byrne, D, and C. Ragin (2009 forthcoming), eds. (2009), Handbook of Case-Centred Research, London: Sage.
- Ragin, C.C. (2008). Redesigning social inquiry: Set relations in social research. Chicago: Chicago University Press.
- Ragin, C. C. (1994). Constructing social research: the unity and diversity of method. Thousand Oaks, CA; London, Pine Forge Press Sage.
- Ragin, C. C. (1997). "Turning the Tables: how case-oriented methods challenge variable oriented methods." Comparative Social Research **16**: 27-42.
- Ragin, C. C. (2000). Fuzzy-set social science. Chicago; London, University of Chicago Press.
- Rihoux, B., & Ragin, C. C. (2009). Configurational comparative methods. Qualitative Comparative Analysis (QCA) and related techniques (Applied Social Research Methods). Thousand Oaks and London: Sage.
- Spencer, J. H. (2008). Household Strategies for Securing Clean Water The Demand for Piped Water in Vietnam's Peri-Urban Settlements. *Journal of Planning Education and Research*, *28*(2), 213-224. doi: 10.1177/0739456x08321793

References II

- Byrne, D. (1999). Complexity theory and the social sciences: an introduction. London; New York, Routledge.
- Byrne, D. (2002). Interpreting quantitative data. London, Sage.
- Byrne, D. (2004). Complex and Contingent Causation: The Implications of Complex Realism for Quantitative Modelling. Making Realism Work: Realist Social Theory and Empirical Research. B. Carter and C. New. London, Routledge.
- Byrne, D. (2005). "Complexity, Configuration and Cases." Theory, Culture and Society 22: 95-111.
- Levi-Faur (2002). Comparative Research Designs in the Study of Regulation: How to Increase the Number of Cases Without Compromising the Strengths of Case-Oriented Analysis. The Politics of Regulation. J. Jacint and D. Levi-Faur. London, Elgar.

References III

- Cooper, B. & Glaesser, J. (2010) Contrasting variable-analytic and case-based approaches to the analysis of survey datasets: exploring how achievement varies by ability across configurations of social class and sex, in Methodological Innovations Online 5(1) 4-23.
- Cooper B. and Glaesser, J. (2008) 'How has educational expansion changed the necessary and sufficient conditions for achieving professional, managerial and technical class positions in Britain? A configurational analysis', Sociological Research Online. Available at: < http://www.socresonline.org.uk/13/3/2.html > 20/04/10
- Ishida, A., Yonetani, M., & Kosaka, K. (2006). Determinants of linguistic human rights movements: an analysis of multiple causation of LHRs movements using a Boolean approach. Social Forces, 84(4), 1937-1955.
- Olsen, W.K. (2009), Non-Nested and Nested Cases in a Socio-Economic Village Study, chapter in D. Byrne and C. Ragin, eds. (2009), Handbook of Case-Centred Research Methods, London: Sage.
- Ragin, C. C. (2008). Qualitative Comparative Analysis using fuzzy sets (fsQCA). In B. Rihoux, & C. C. Ragin (eds), Configurational comparative methods. Qualitative Comparative Analysis (QCA) and related techniques . Thousand Oaks and London: Sage.
- Rihoux, B. (2006). Qualitative Comparative Analysis (QCA) and related systematic comparative methods: recent advances and remaining challenges for social science research. International Sociology, 21(5), 679-706.
- Rihoux, B., & Grimm, H. (2006). Innovative Comparative Methods for Policy Analysis. Beyond the Quantitative-Qualitative Divide. New York: Springer/Kluwer.
- Rohwer, Gotz (2010). Qualitative Comparative Analysis: A Discussion of Interpretations. European Sociological Review.. DOI: 10.1093/esr/jcq034