Qualitative Comparative Analysis

Computational Social Science

Lecture 4

Overview

QCA is a systematic way to compare many cases that have an outcome of interest

There are three forms of QCA that differ in the way they calibrate case attributes

- csQCA (crisp set QCA) dichotomises all attributes, to true or false
- mvQCA (multi-value QCA) allows multiple discrete values for attributes
- fsQCA (fuzzy set QCA) allows subjective continuous values in [0,1]

All versions use set theory to link outcomes to patterns of attributes

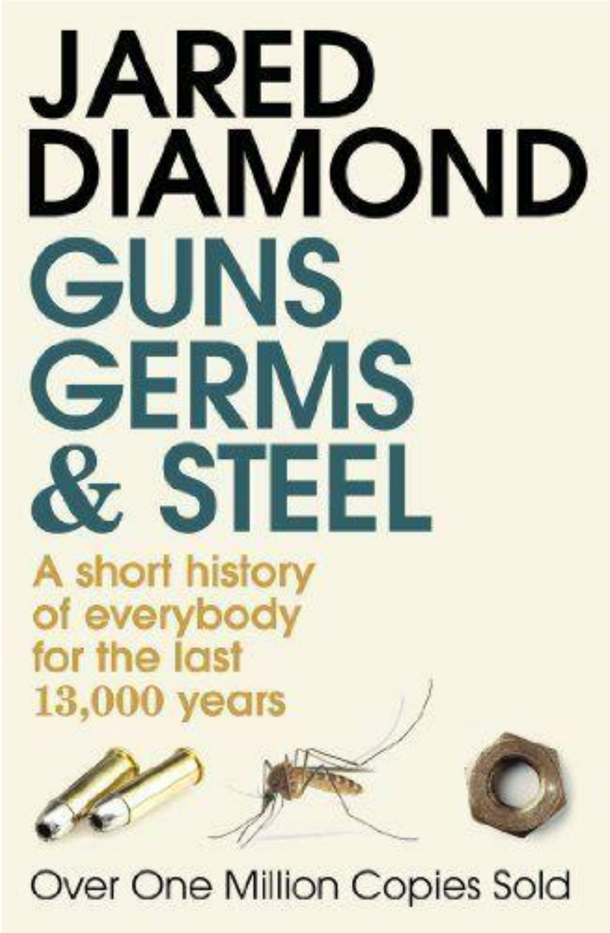
- Attribute combinations (referred to as causal recipes) are derived from logical, rather than probabilistic, foundations

Orientation: Guns, Germs and Steel

Diamond seeks to explain the differential progress of civilisations and source of global inequality

"Why is it that you white people developed so much cargo and brought it to New Guinea, but we black people had little cargo of our own?"

- Question by Yali, a political activist in Papua New Guinea during one of Diamond's field research visits (ornithology, cultural anthropology)



Durham University

Diamond’s explanation

Europeans had access to large scale agriculture

- Plants that could become crops
- Animals suitable for meat and labour that could be domesticated

Agriculture is necessary for specialisation, and hence innovation

East-West orientation promoted migration because the same climate reduced need for adaptation

- Spreads communication, innovations

Dense population and close living with animals exposed Europeans to a wide range of disease, so broad immunity

Geography (mountains, rivers) promoted separate communities that fought and competed so a range of opportunities

How would you develop and test such a theory?

Theory is that many conditions contribute

Case study research

- Cases are different societies
- Many cases needed

Identify causally relevant conditions that are present for societies that flourish

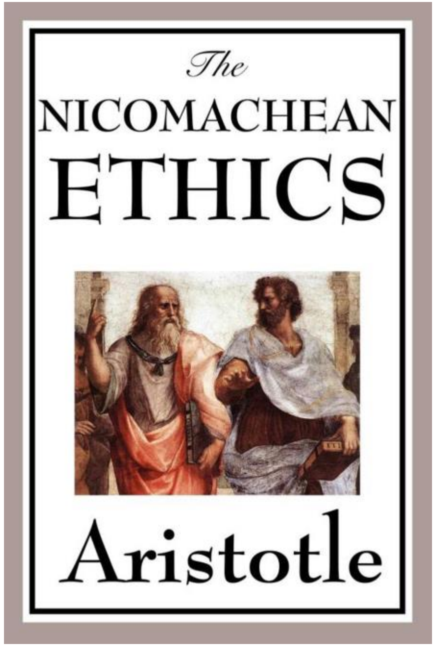
- Always present? – suggests necessary

Check whether those conditions are present for societies that do not flourish

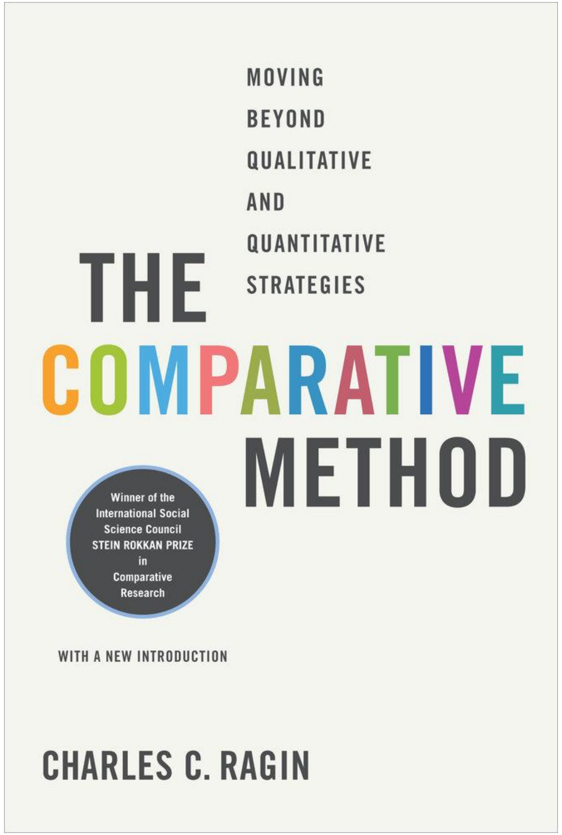
- If so, irrelevant or at least not sufficient

One may go wrong in many different ways … but right only in one

- Book II, Ch V



Configurational complexity



QCA was developed by Charles Ragin to analyse questions in the form: Why does a social phenomenon occur in some situations but not others?

It is a structured method for case based comparative research

- Select attributes that appear to be relevant
- Compare attribute patterns where the phenomenon occurs and where it doesn’t

Durham University

Logic foundations

Input Propositions		Expression with P and Q
P	Q	
T	T	Either T or F
T	F	Either T or F
F	T	Either T or F
F	F	Either T or F

Durham University

QCA uses the language of logic and set theory

QCA uses several forms of conditionals that have precise definitions and meaning

- Necessary
- Sufficient
- INUS (insufficient but necessary part of an unnecessary but sufficient condition)

Cases are presented in the form of truth tables

Formal logic and set theory are different ways of representing the same concepts

Lots of terminology here, ask questions if you get lost

Structure of Boolean algebra

Algebra is a mathematical term for a system of symbols and rules to manipulate them

In Boolean algebra, propositions are the symbols

The important operators are

- Negation (NOT)
- Conjunction (AND)
- Disjunction (OR)

Operators

- Combine propositions into more complicated expressions
- Set out rules for calculating the truth value of the constructed expression

They are the logic equivalent of operators such as multiple, add, subtract in the algebra of numbers

What is a proposition?

A proposition is a statement that has a truth value

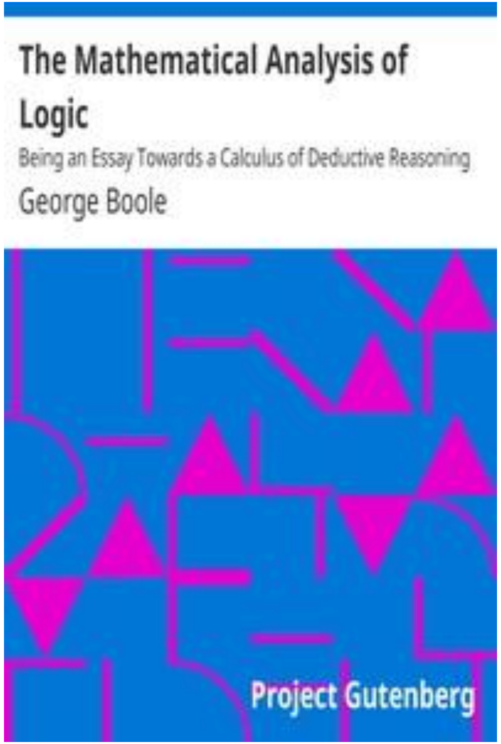
- In binary (Boolean) logic, truth values is TRUE or FALSE

Examples

- Squares have 4 sides (TRUE)
- The cat is alive (value depends on the specific cat)

In formal logic, propositions are given letters like A, B, P or Q

- For example, P might stand in for “Squares have 4 sides”



The basics of logic were formalised by Boole in the mid 1850s. Variables that can only take the values T or F are referred to as Boolean variables.

What if the attribute is not binary?

Many social science statements are not clearly T or F

Which of these countries are (or are not) democracies?

- North Korea: elections with one candidate
- Belarus: opposition arrested, claims of widespread rigged counting
- Thailand: military coup in 2014, elections in 2017 for half of parliament with other half of members appointed by junta
- USA: can lose despite >50% of votes due to boundaries and State votes
- UK: 67% voter turnout, Prime Minister selected by party

In formal logic, each of these countries are either democracies or not, no ‘sort of’

- Fuzzy logic allows partial truth, later in lecture

Truth tables

Tool to set out the truth value of a statement for all potential combinations of truth values for the inputs

Number of rows depends on the number of input propositions

Multiplies by 2 for each additional

- 2 rows with 1 input, 4 for 2 inputs, 8 for 3 inputs etc

Input Propositions		Expression with P and Q
P	Q	
T	T	Either T or F
T	F	Either T or F
F	T	Either T or F
F	F	Either T or F

Typical form of truth table for two variables

Negation (Logical NOT)

Negation means “it is not the case that…”

Symbol is $\neg$

P	$\neg P$
T	F
F	T

Example

- 5 > 3 is TRUE so $\neg$ (5 > 3) is FALSE

Other notation for $\neg P$

- $\sim P$
- !P
  - Notably for computer languages

Conjunction (Logical AND)

Conjunction means “both … and …”

Symbol is $\wedge$

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

Example

- B = “5 > 3” is TRUE
- S_1 = “Squares have 4 sides” is TRUE
- S_2 = “Squares have 5 sides” is FALSE
- So $B \wedge S_1$ is TRUE and $B \wedge S_2$ is FALSE

Other notation for $P \wedge Q$

- $P \,\&\, Q$
- $P * Q$
- $P \cdot Q$
- PQ

Disjunction (Logical OR)

Disjunction means “either … or …”

Symbol is $\vee$

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

Example

- B = “5 > 3” is TRUE
- S_2 = “Squares have 5 sides” is FALSE
- So $B \vee S_2$ is TRUE

Other notation for $P \vee Q$

- P + Q

Expressions use combinations of these (and other) operators

$(\neg A \vee B) \wedge \neg(B \wedge C)$

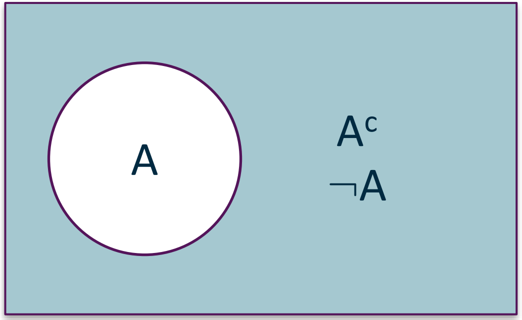
A	B	C	$\neg A$	$\vee B$	$\wedge$	$\neg$	$B \wedge C$
T	T	T					
T	T	F					
T	F	T					
T	F	F					
F	T	T					
F	T	F					
F	F	T					
F	F	F					
Calculate in this order:			1	2	5	4	3

Expressions use combinations of these (and other) operators

$(\neg A \vee B) \wedge \neg(B \wedge C)$

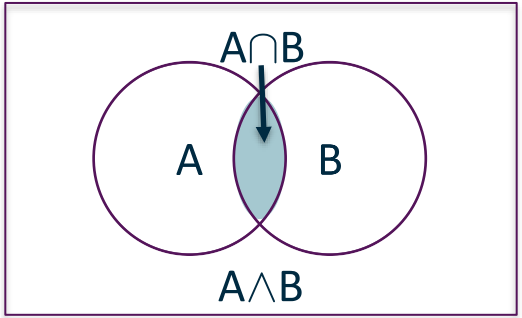
A	B	C	$\neg A$	$\vee B$	$\wedge$	$\neg$	$B \wedge C$
T	T	T	F	T	F	F	T
T	T	F	F	T	T	T	F
T	F	T	F	F	F	T	F
T	F	F	F	F	F	T	F
F	T	T	T	T	F	F	T
F	T	F	T	T	T	T	F
F	F	T	T	T	T	T	F
F	F	F	T	T	T	T	F
		Calculate in this order:	1	2	5	4	3

Conceptualising as sets of cases: attribute present set to TRUE



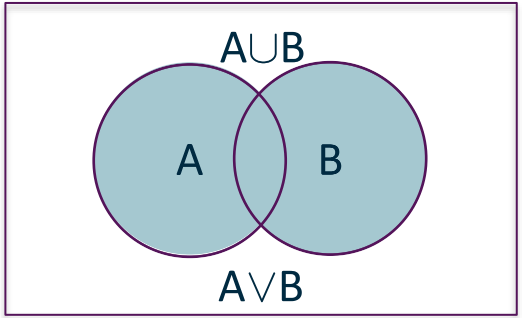
A is the set of cases where attribute A is present

Logical NOT = set complement



A, B are the sets of cases where attributes A, B are present

Logical AND = set intersection



A, B are the sets of cases where attributes A, B are present

Logical OR = set union

We want to make causal claims

https://commons.wikimedia.org/wiki/File:Billard.JPG

David Hume (Scottish philosopher, 1700s)

- Strict empiricist, knowledge from perception
- Argued that causation cannot be observed
  - When you hit a billiard ball into another, the only way you know if and how the second ball will move is from experience
- We see only that two events follow each other

Assert causal relationships by observing *constant conjunctions*

- C causes E means if an occurrence of C is always followed by an occurrence of E
- Regular (ie always) association from C to E

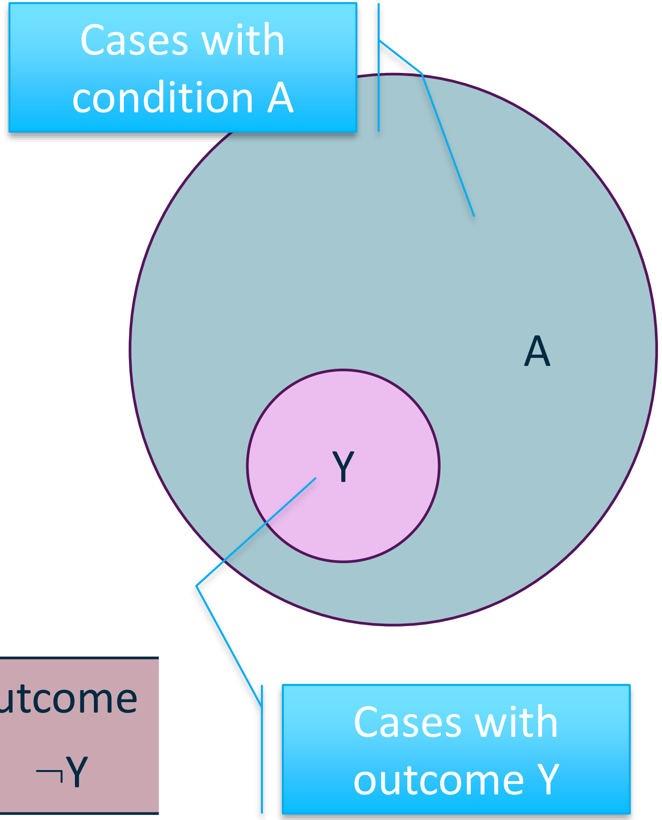
Durham University

Necessary conditions

A is necessary for outcome Y if Y cannot come about without A

- All cases with Y also have A
- No cases have Y but not A

Necessity says nothing about whether Y actually occurs following A



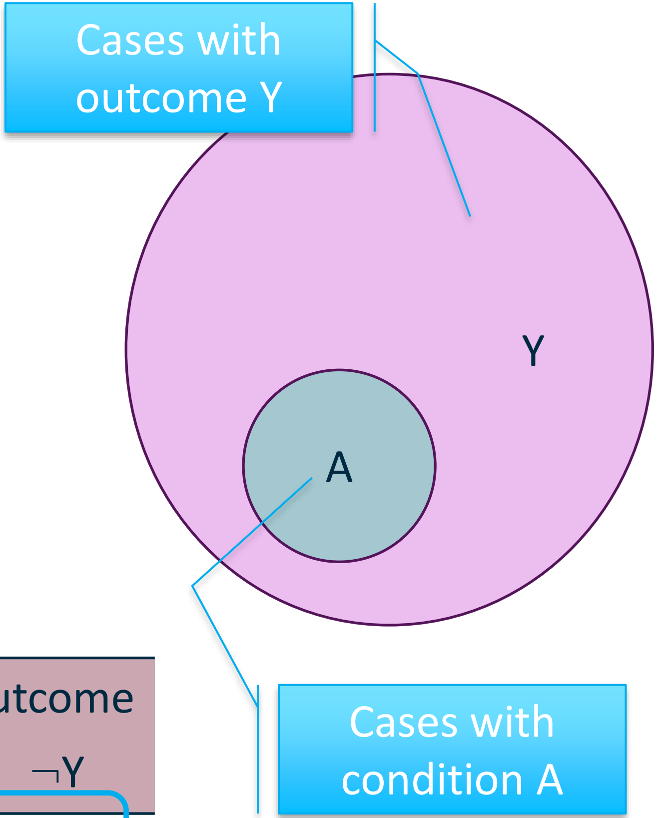
Conditions		Cases by outcome	
A	B	Y	$\neg$Y
T	-	$\geqslant$1	any
F	-	0	any

Sufficient conditions

A is sufficient for outcome Y if A cannot happen without Y also happening

- All cases of A also have Y

Sufficiency says nothing about whether Y could occur in ways without A



Conditions		Cases by outcome	
A	B	Y	$\neg$Y
T	-	$\geqslant 1$	0
F	-	any	any

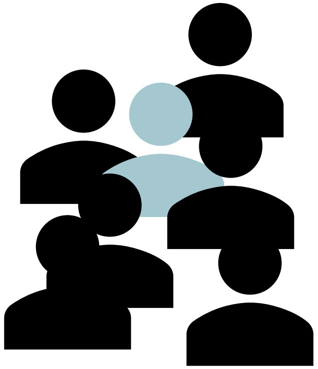
Refining causal claims

John Stuart Mill (English philosopher, 1800s) noted that differences in patterns refine causal claims

Method of agreement: *If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance … is the cause (or effect)*

Method of difference: *If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, … the circumstance … is the effect, or cause,* ***or an indispensable part of the cause****, of the phenomenon*

Mill also set out three further methods that built on these two



JS Mill (1843). A System of Logic

INUS: Insufficient but necessary part of sufficient combination

Outcomes typically arise from combination of conditions

- Single conditions insufficient
- Individual conditions that contribute to combination are referred to as “an **I**nsufficient but **N**ecessary part of a condition which is itself **U**nnecessary but **S**ufficient for the result”

Example

- Short circuit and flammable material sufficient for fire (but not necessary as there are other ways in which fires can start)
  - Actually, more conditions such as oxygen
- Short circuit is INUS causal but cannot start fire independently

JL Mackie (1965). Causes and conditions
https://www.jstor.org/stable/20009173

INUS: Insufficient but necessary part of sufficient combination

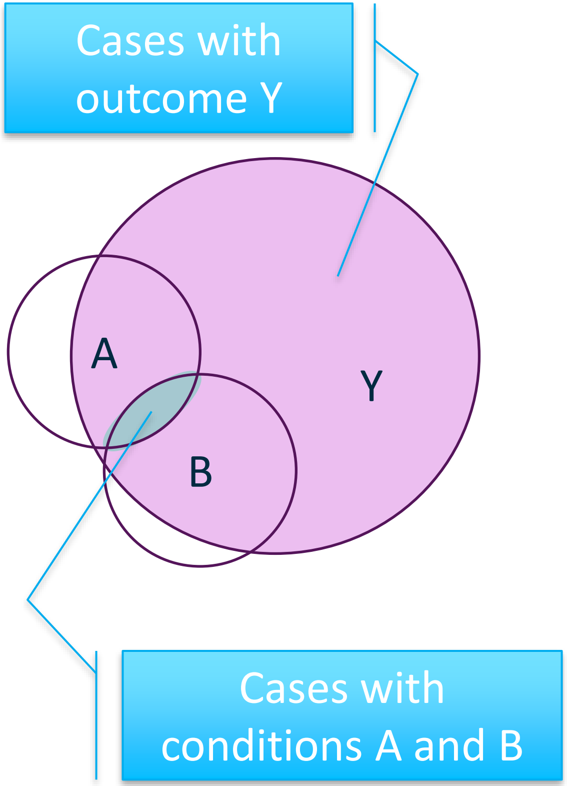
A and B are both insufficient for Y by themselves

- Some cases with A, B outside of Y set

A and B are both INUS for Y

- Intersection of A and B is inside Y set

Conditions		Cases by outcome	
A	B	Y	¬Y
T	T	$\geqslant 1$	0
T	F	any	$\geqslant 1$
F	T	any	$\geqslant 1$
F	F	any	any



Application to cases

How is this useful for analysing cases?

Describing and analysing theory

In QCA, propositions concern the presence/absence of attributes

QCA is a systematic analysis of multiple cases that are diverse on both attributes and outcome

The goal is to find combinations of attribute presence/absence that

- Occur for cases with some outcome of interest
- Do not occur for cases without the outcome of interest

Language of QCA

Conditions (attributes) phrased with

- present / TRUE encourages outcome
- absent / FALSE suppresses outcome

Configurations

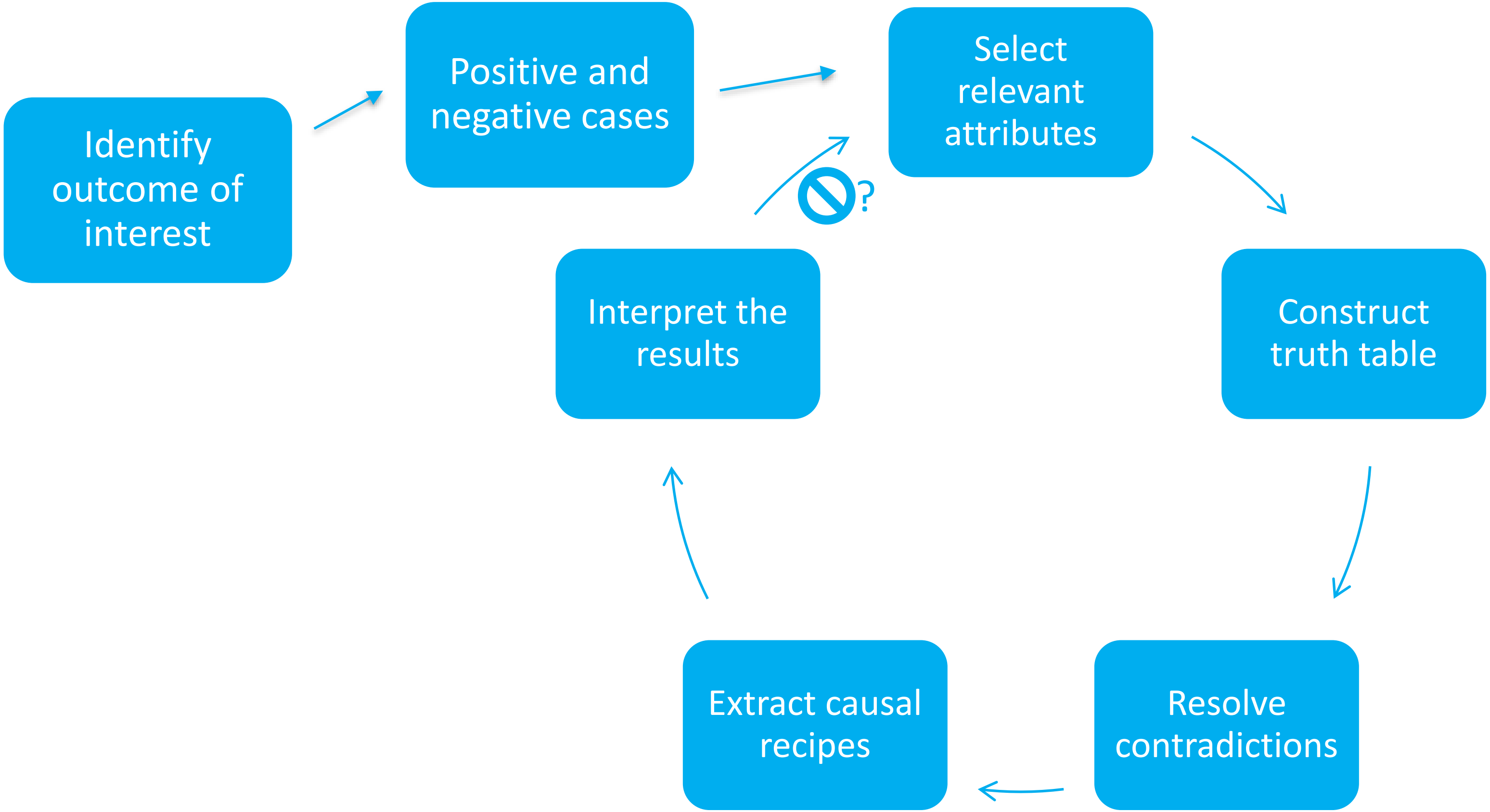
- Combinations of conditions

Causal recipe

- Logical expression for a configuration

By convention, lower case for NOT (absent)

QCA phases



Durham University

Outcome of interest: What is the social phenomenon that you are seeking to understand?

Argentina anti-austerity strike brings country to standstill (25 June 2018), BBC News
https://www.bbc.co.uk/news/world-latin-america-44607051

Example with **hypothetical data**

Mass protests in response to IMF mandated austerity measures as conditions for debt renegotiation

Durham University

Diverse cases, plausible relevant conditions

Positive cases

- Austerity protests have occurred
- Peru, Argentina, Tunisia…

Negative cases, similar experiences

- No protests despite IMF austerity
- Mexico, Costa Rica…

Inspiration:
Walton and Ragin (1990), Global and National Sources of Political Protest: Third World Responses to the Debt Crisis
https://doi.org/10.2307/2095752

Relevant conditions to include in analysis

- Austerity severity
- Prior political mobilisation
- Living conditions
- Price increases
- Government corruption
- Economic dependence

Simplify by combining similar conditions

Durham University

Cases need several levels of diversity

QCA derives its explanatory power by comparing cases

- Cases with different attribute values add information more than many identical cases
- Condition combinations with no cases are ***remainders***

Diversity of cases involves

- Positive and negative outcomes
- For each condition, prefer at least one third of cases have present and at least one third absent
- Check that condition distributions are different (that is, no pair of conditions have similar patterns of outcome)

Rihoux & Ragin (2009) Box 3.5 (pg 45)

Construct extended truth table

Either all conditions (attributes) or a subset of conditions that are of particular interest

Truth table has

- Rows for present/absent each condition
- If *n* conditions, truth table has 2^n rows

Two additional columns

- Outcome type (occurs, does not occur, contradiction)
- Number of cases

Consistency score

- For each condition combination (row)
- Proportion of cases with outcome
- Consistent if 0 or 1

Contradiction

- Same conditions, different outcomes

Initial truth table for (hypothetical) austerity protest example

Prior protests?	Severe austerity?	Govt corrupt?	Rapid price rise?	Cases: protest	Cases: not protest
Yes	Yes	Yes	Yes	8	0
Yes	Yes	Yes	No	6	2
Yes	Yes	No	Yes	6	0
Yes	Yes	No	No	1	5
Yes	No	Yes	Yes	0	0
Yes	No	Yes	No	0	10
Yes	No	No	Yes	1	7
Yes	No	No	No	0	3
No	Yes	Yes	Yes	5	0
No	Yes	Yes	No	0	0
No	Yes	No	Yes	4	0
No	Yes	No	No	0	0
No	No	Yes	Yes	1	5
No	No	Yes	No	0	4
No	No	No	Yes	0	0
No	No	No	No	0	0
				32	36

Durham University

Ragin (2008). What is Qualitative Comparative Analysis (QCA)? Workshop slides
https://eprints.ncrm.ac.uk/id/eprint/250/1/What_is_QCA.pdf

Identify contradictions

Prior protests?	Severe austerity?	Govt corrupt?	Rapid price rise?	Cases: protest	Cases: not protest
Yes	Yes	Yes	Yes	8	0
Yes	Yes	Yes	No	6	2
Yes	Yes	No	Yes	6	0
Yes	Yes	No	No	1	5
Yes	No	Yes	Yes	0	0
Yes	No	Yes	No	0	10
Yes	No	No	Yes	1	7
Yes	No	No	No	0	3
No	Yes	Yes	Yes	5	0
No	Yes	Yes	No	0	0
No	Yes	No	Yes	4	0
No	Yes	No	No	0	0
No	No	Yes	Yes	1	5
No	No	Yes	No	0	4
No	No	No	Yes	0	0
No	No	No	No	0	0

Durham University

Dealing with contradictions

Contradictions must be resolved

- Consistent outcome for each row

Modify the conditions

- More conditions, that separate the contradictory cases
  - But many remainders
- Revise or replace some conditions
- Reconsider present/absent threshold

Modify the cases

- Narrow the outcome definition
- Consider whether any of the cases are out of scope

Ad hoc adjustments

- Set all cases in contradictory rows to outcome did not occur
- Set minority outcome to majority

Rihoux & Ragin (2009) Box 3.6 (pp 48-49)

Hypothetical contradiction corrections

Re-examination of the cases without protests showed that they had severely repressive regimes

- Add not-repressive to the conditions
- Assess all cases for that condition

Re-examination of the three rows where there was a single example of a protest showed that these were not countries where a protest emerged, but instead it spread from a neighbouring country

- Remove those no-protest cases

Cases: protest	Cases: not protest
8	0
6	2
6	0
1	5
0	0
0	10
1	7
0	3
5	0
0	0
4	0
0	0
1	5
0	4
0	0
0	0

Consistent truth table
(20 remainder rows not shown)

Prior protests?	Severe austerity?	Govt corrupt?	Rapid price rise?	Not repressive?	Cases: protest	Cases: not protest
Yes	Yes	Yes	Yes	Yes	8	0
Yes	Yes	Yes	No	Yes	6	0
Yes	Yes	Yes	No	No	0	2
Yes	Yes	No	Yes	No	6	0
Yes	Yes	No	No	Yes	0	5
Yes	No	Yes	No	No	0	10
Yes	No	No	Yes	Yes	0	7
Yes	No	No	No	No	0	3
No	Yes	Yes	Yes	Yes	5	0
No	Yes	No	Yes	No	4	0
No	No	Yes	Yes	No	0	5
No	No	Yes	No	No	0	4

Durham
University

Summary by condition

Severe austerity is necessary for protests to erupt

- Protests do not occur unless the austerity is severe
- But not sufficient – protests can occur without severe austerity

Other attributes show no specific relationship

Prior	Protests	Not Protest
Yes	20	27
No	9	9

Austerity	Protests	Not Protest
Yes	29	7
No	0	29

Corrupt	Protests	Not Protest
Yes	19	21
No	10	15

Rapid Rise	Protests	Not Protest
Yes	23	12
No	6	24

Not repress	Protests	Not Protest
Yes	19	12
No	10	24

Extract causal recipes (Quine McCluskey Method)

The truth table provides recipes directly

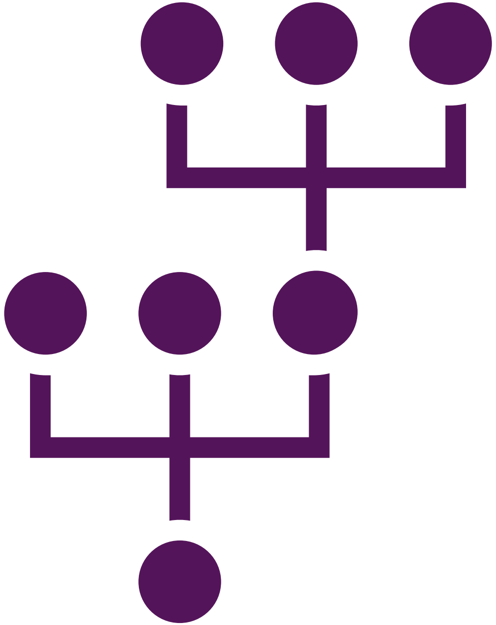
- Each row is a recipe as there is a consistent outcome

Extracting the recipes is simplifying this overfitted solution

- Iteratively combine pairs of rules
- Pair differs on only one attribute

In practice, done with software

Recipes for outcome occurs are separate from recipes for outcome does not occur



Focus on rows with positive (outcome occurred) cases

P Prior protests?	A Severe austerity?	C Govt corrupt?	R Rapid price rise?	N Not repressive?	Cases: protest	Cases: not protest
Yes	Yes	Yes	Yes	Yes	8	0
Yes	Yes	Yes	No	Yes	6	0
Yes	Yes	No	Yes	No	6	0
No	Yes	Yes	Yes	Yes	5	0
No	Yes	No	Yes	No	4	0

Pair with only single attribute different: Rapid price rise (R)

- All protests for $P \wedge A \wedge C \wedge R \wedge N$
- All protests for $P \wedge A \wedge C \wedge \neg R \wedge N$
- Simplify to $P \wedge A \wedge C \wedge N$
  - That is, the truth value of R does not matter

In QCA notation: PACN

Can also consider remainder (no cases) rows

P Prior protests?	A Severe austerity?	C Govt corrupt?	R Rapid price rise?	N Not repressive?	Cases: protest	Cases: not protest
Yes	Yes	Yes	-	Yes	14	0
Yes	Yes	No	Yes	No	6	0
No	Yes	Yes	Yes	Yes	5	0
No	Yes	Yes	Yes	No	0	0
No	Yes	No	Yes	No	4	0

Same table as before but one remainder row shown

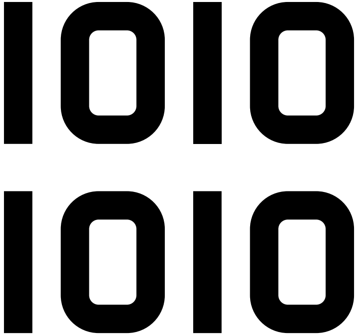
Pair is row with cases and remainder row

- Row with cases must have an absent condition (No, FALSE)
- Remainder row must have same condition present (Yes, TRUE)

On theoretical grounds, any present attribute (P, A, C, R, N) increases the likelihood of protests occurring, so remainder interpreted as if any cases would have had outcome

Simplify to: $\neg P \wedge A \wedge R \wedge \neg N$, written as pARn

Various levels of solutions



Software returns multiple solutions that reflect different stringency approaches

Complex solution does not use remainder configurations

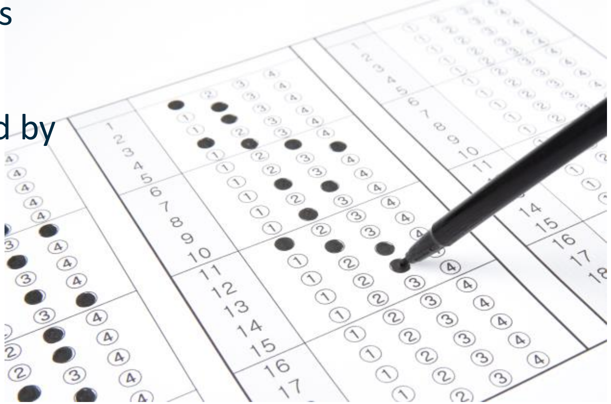
- Has the most terms
- For this example, would be AcRn + PACN + SCRN

Parsimonious solution, deletes remainder rows

- Assumes remainders would not contradict any identified configurations
- For this example, would be CN + AR

Intermediate solution allows remainder rows that are theoretically justified by analyst

- For this example, would be PACN + AR



Interpret the results: how to use the recipes

Consider further the intermediate solution PACN + AR

- With factoring, A(PCN + R)

This means: Protests erupt if severe austerity occurs with either rapid price rises or with (all of) prior protests and government corruption and a not repressive regime

$A \wedge R$ is sufficient

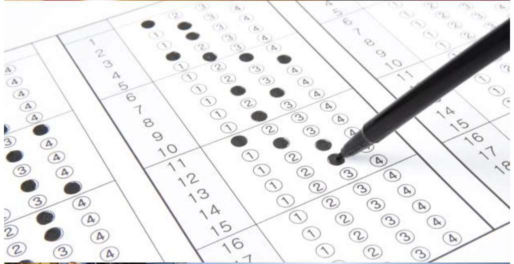
- Visible in two condition truth table

Could also show for $A \wedge P \wedge C \wedge N$

Conditions		Protests	
A	R	Yes	No
T	T	23	0
T	F	6	7
F	T	0	12
F	F	0	17

Durham University

Interpret the results: what do the recipes tell you?



With real data, connect the recipes back to cases and theory

Do the recipes make sense?

How many (and which) cases relate to each part of the recipe?

- Do these groupings provide further insight?

Are there cases that were not considered that are similar to the developed recipes?

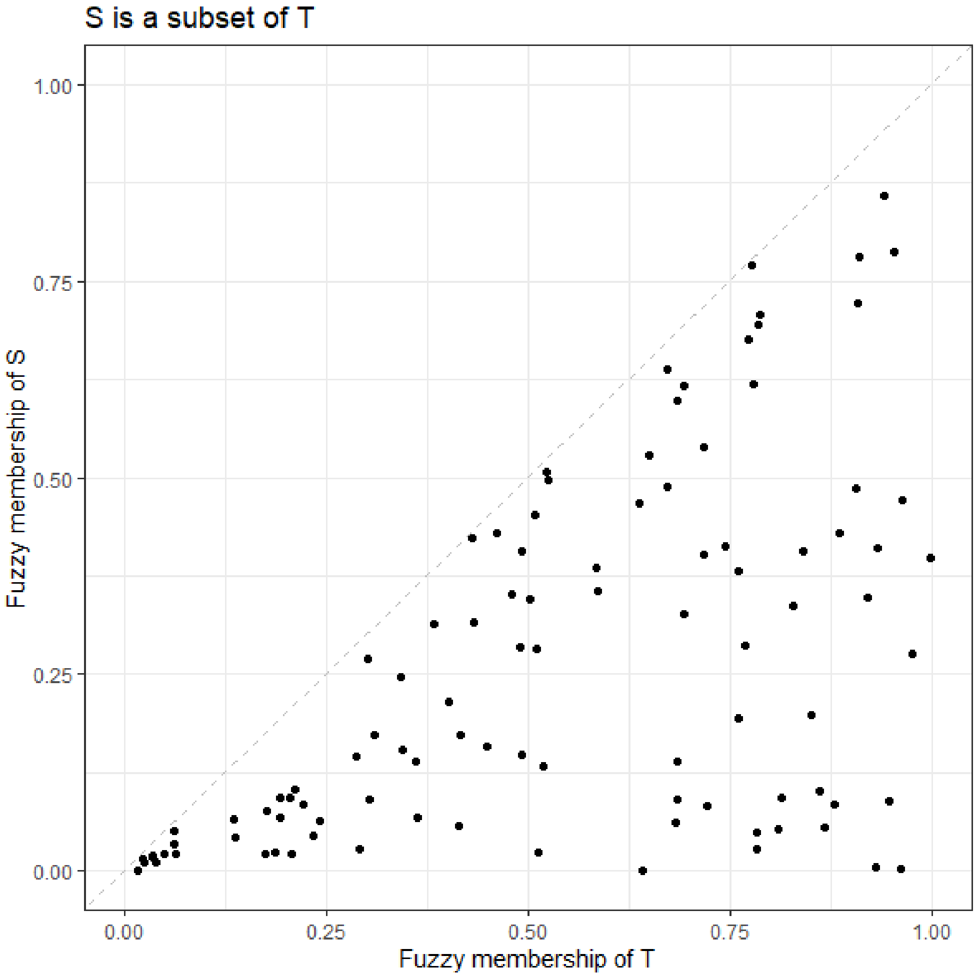
Questions for discussion

How does QCA deal with complexity?

In what ways are clustering and QCA similar and different?

Like regression, QCA constructs a model that explains an outcome with multiple input features. How do the models differ?

Fuzzy set theory



Durham University

Operationalising attributes to discrete values

Clear recipe extraction with yes/no attributes

This requires some boundary for each attribute

- Scored highly: summative score of 70 or higher
- Prior protests: at least 100 people on at least three occasions
- Rapid price rises: inflation more than 50% in one month
- Animal domestication: can be used for labour and 200+ kg

Divisions should be justifiable on theoretical, substantive or technical grounds rather than set arbitrarily

csQCA and mvQCA both measure in discrete values

Heap

https://en.wikipedia.org/wiki/Sorites_paradox#/media/File:Tannin_heap.jpeg

Not Heap

Binary classification is a major criticism of crisp-set QCA

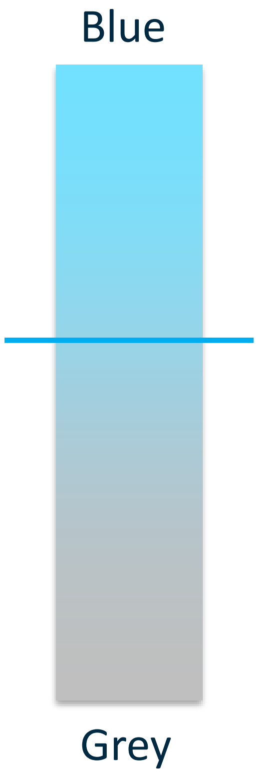
If we are concerned with complexity, does it make sense to simplify the data we are working with so drastically?

- Classification erases detail about our cases

One response: Complexity emphasises differences in kind, not degree

- Care about qualitative differences between cases
- Don’t care about minor, quantitative differences

Nevertheless, many relevant attributes vary gradually and are difficult to classify



Fuzzy-Set membership varies from 0 (out) to 1 (in)

These are really multiple discrete, not fuzzy

Crisp-Set (2 values)	**3-value fuzzy set**	**5-value fuzzy-set**	**7-value fuzzy set**	**Continuous fuzzy set**
1 = fully in **0** = fully out	**1** = fully in **0.5** = not fully in or out **0** = fully out	**1** = fully in **0.75** = more in than out **0.5** = not fully in or out **0.25** = more out than in **0** = fully out	**1** = fully in **0.83** = mostly but not fully in **0.67** = more or less in **0.5** not fully in or out **0.33** = more or less out **0.87** = mostly but not fully out **0** = fully out	**1** = fully in **Numerical scores** indicating more in than out **0.5** not fully in or out **Numerical scores** indicating more out than in **0** = fully out

Durham University

Ragin (2009). Redesigning Social Inquiry : Fuzzy Sets and Beyond (pg 31)

Moving to fuzzy attributes raises complications

Need rules to set attribute values consistently

- Referred to as calibration

How do set operations work?

- What does a subset mean?
- What is necessary or sufficient?

How do truth tables work?

- Which configuration does a case belong to?
- What is the case's membership value for that configuration?
- How to report the recipes?

Some aspects of these changed 2000 to 2008 so use recent references

Durham University

Calibration involves judgement

Fuzzy set membership values are conceptual, they should highlight meaningful differences

No case should have any value at exactly 0.5

- Technical: cannot allocate to truth table

For numerical attributes

- Use anchor points
- Function for intermediate values

For descriptive attributes

- Need to compare (order) cases
- Assign membership individually

Example, numerical: regular attendance

- Value for 9 out of 10 sessions?
- Value for 6 out of 10 sessions?

Example, descriptive: democracy

- Value for Thailand?
- Value for United Kingdom?

Calibrating numerical values

$$dm_x = \begin{cases} 0 & \text{if } x \le e, \\ \frac{1}{2}\left(\frac{e-x}{e-c}\right)^b & \text{if } e < x \le c, \\ 1 - \frac{1}{2}\left(\frac{i-x}{i-c}\right)^a & \text{if } c < x \le i, \\ 1 & \text{if } x > i. \end{cases}$$

e is exclusion (0) anchor
c is crossover (0.5) anchor
i is inclusion (1) anchor

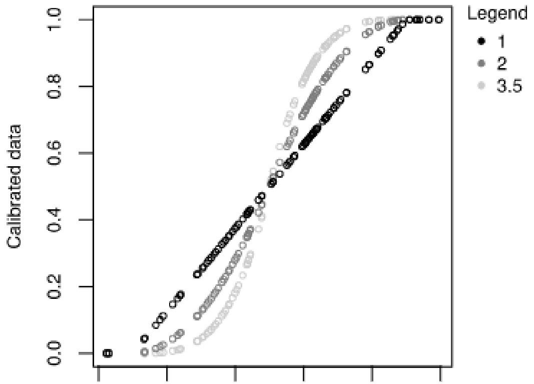
For quantitative attributes, fuzzy set membership values use rescaling

Fixed value anchors at 0, 0.5 and 1

- 3 anchors if ideal is extreme ('tall')
- 6 anchors if ideal is middle ('average height')

Function to transform the values between the anchor points

- Logistic and power functions are common



Duşa (2018), QCR with R, s4.2

Subset of a fuzzy set

U is universal set with members u_j

S and T are two fuzzy sets

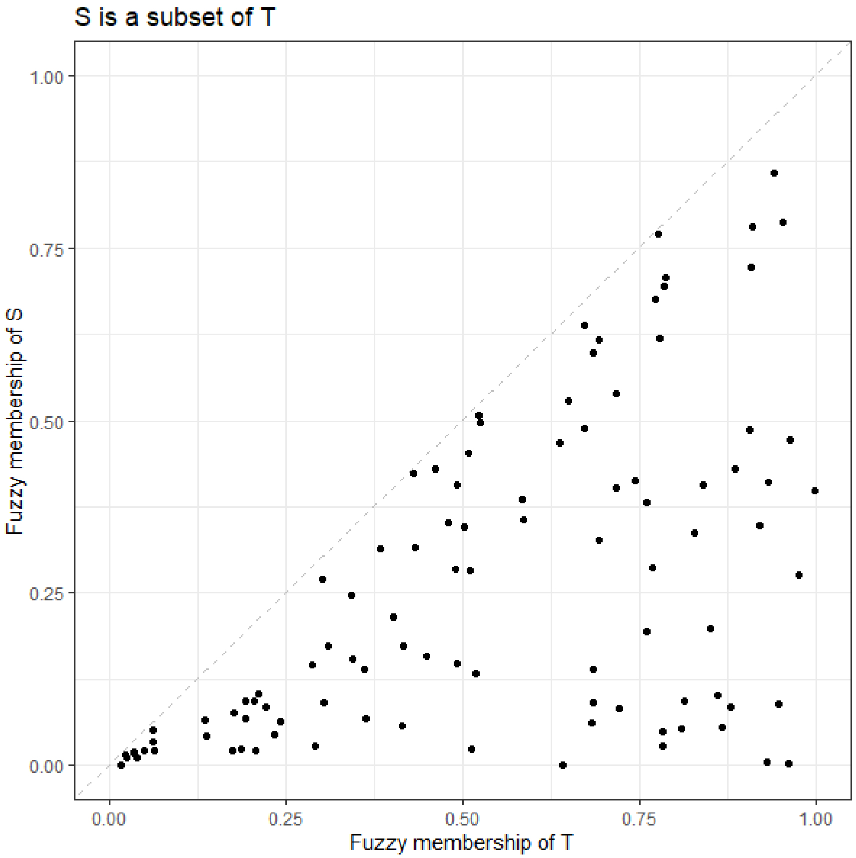
S is a subset of T (notated $S \subseteq T$) means

- membership values in S are less than or equal the values in T, all members

$$S(u_j) \leqslant T(u_j) \qquad \forall\, u_j \in U$$

Scatterplot of the two membership values gives characteristic triangular plot

- Note that plot would be upper left triangle for T is subset of S



Durham University

Fuzzy set operators

Negation (NOT, complement) is the 'opposite' value (1 – *a* if *a* is the set membership)

- Note that 'NOT tall' is not the same as 'short'

Conjunction (AND, intersection) is minimum, the largest fuzzy set contained in both

Disjunction (OR, union) is maximum, the smallest fuzzy set that contains both

These operators are the same as for crisp sets with 0 for FALSE and 1 for TRUE

A	B	$\neg$A	A$\wedge$B	A$\vee$B
a	b	1-a	min(a,b)	max(a,b)
0.3	0.6			
0	1			

Fuzzy set operators

Negation (NOT, complement) is the 'opposite' value ($1 - a$ if a is the set membership)

- Note that 'NOT tall' is not the same as 'short'

Conjunction (AND, intersection) is minimum, the largest fuzzy set contained in both

Disjunction (OR, union) is maximum, the smallest fuzzy set that contains both

These operators are the same as for crisp sets with 0 for FALSE and 1 for TRUE

A	B	$\neg$A	A$\wedge$B	A$\vee$B
a	b	1-a	min(a,b)	max(a,b)
0.3	0.6	0.7	0.3	0.7
0	1	1	0	1

Necessity and sufficiency in fuzzy sets

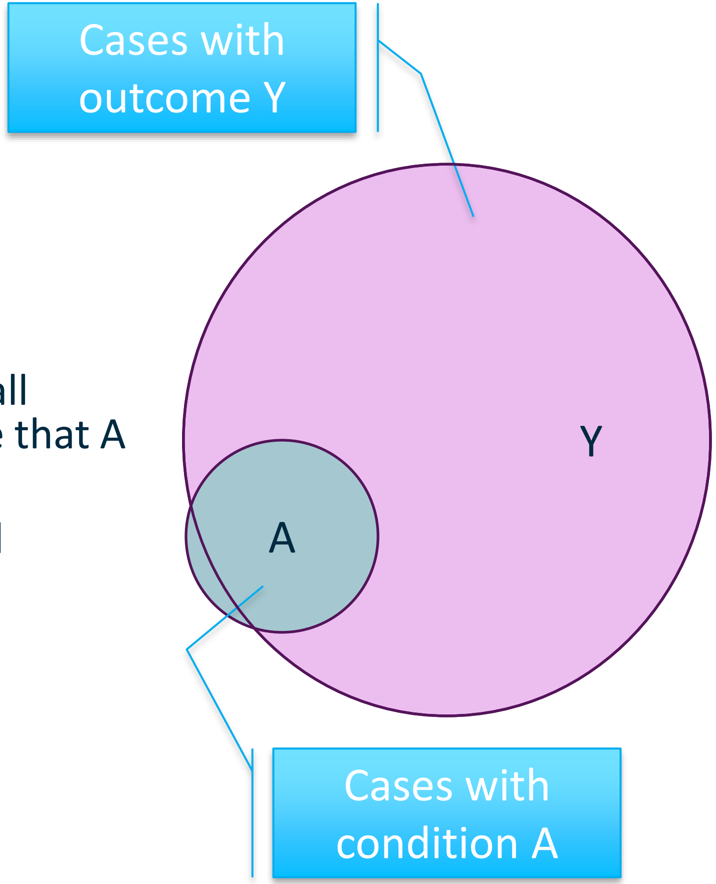
For condition A and outcome Y

- A is sufficient for Y: $A(u_j) \leqslant Y(u_j)$
- A is necessary for Y: $Y(u_j) \leqslant A(u_j)$

With fuzzy sets, can also have 'almost' subsets

- If almost all the cases satisfy the inequality and a small number 'almost' do, there is still substantial evidence that A is sufficient for Y
- For $\leqslant$ scatterplot: a few cases 'just' over the diagonal

Inclusion of A in Y is calculated as (A∧Y)/A = $\frac{\sum_j min(a_j, y_j)}{\sum_j a_j}$



Durham University

Truth tables with fuzzy sets

```
 OUT: output value
   n: number of cases in configuration
incl: sufficiency inclusion score
 PRI: proportional reduction in inconsistency
```

	A	I	M	U	OUT	n	incl	PRI	cases
6	0	1	0	1	0	1	0.760	0.400	IT
8	0	1	1	1	1	1	0.870	0.667	IE
9	1	0	0	0	1	1	1.000	1.000	NL
10	1	0	0	1	0	2	0.700	0.438	BE,DK
12	1	0	1	1	0	2	0.536	0.071	NO,SE
13	1	1	0	0	1	3	0.971	0.944	FR,DE,US
14	1	1	0	1	1	1	0.821	0.583	AU
16	1	1	1	1	0	1	0.654	0.100	UK

As fuzzy membership can take any value, a truth table with each value would have very many rows and each case would likely occupy a separate row

Instead, cases are allocated to the closest 'corner'

- Set to 1 if membership > 0.5 and to 0 if < 0.5
- Example: (0.3, 0.2, 0.7) $\rightarrowtail$ (0, 0, 1) $\rightarrowtail$ FFT

As well as the cases allocated, the truth table reports the inclusion value for that set of cases

- Measure of sufficiency of attributes for outcome

Duşa (2018), QCR with R, pg 151

Durham University

QCA

(wrap up)

Durham University

QCA bridges qualitative and quantitative

Qualitative

Deep familiarity with the cases required for the iterations between theory, relevant attributes, and evidence

Moderate number of cases

- More than can be analysed manually
- Fewer than for statistical methods

Quantitative

Identification of cross-case patterns

Use of mathematical elements

- Calibration of attribute values
- Boolean (crisp) and fuzzy logic operators

In what ways is QCA a case based complexity method?

Case is a complex system

- Cases are treated as a composite of their attributes

Case is the unit for developing recipes

- All attributes are used to assess sufficiency and necessity
- Recipes respect diversity and heterogeneity of the cases and their contexts

Recipes have a close relationship with theory, so cases are examples of application of theory

```
 OUT: output value
   n: number of cases in configuration
incl: sufficiency inclusion score
 PRI: proportional reduction in inconsistency
```

	A	I	M	U	OUT	n	incl	PRI	cases
6	0	1	0	1	0	1	0.760	0.400	IT
8	0	1	1	1	1	1	0.870	0.667	IE
9	1	0	0	0	1	1	1.000	1.000	NL
10	1	0	0	1	0	2	0.700	0.438	BE,DK
12	1	0	1	1	0	2	0.536	0.071	NO,SE
13	1	1	0	0	1	3	0.971	0.944	FR,DE,US
14	1	1	0	1	1	1	0.821	0.583	AU
16	1	1	1	1	0	1	0.654	0.100	UK

Attributes are configurational rather than additive

Regression models attempt to explain the relationship between both large and small values of the same pair of attributes in the same model

- Differences in values add to the outcome

QCA separates the large and the small values, including them in different configurations

- Concerned with meaningful differences of kind rather than small differences in values

	Cases by outcome	
A	Y	¬Y
T	⩾1	0
F	any	any

Irrelevant to sufficiency
Conflicts with regression

Irrelevant to sufficiency
Supports regression

Durham University

Two significant challenges

Diverse cases are essential

- The logic of comparing sets of truth values relies on many combinations available from the dataset
- Too many remainders limit the opportunity to identify recipes

Operationalising attributes (crisp or fuzzy) relies on judgement and iteration

Next session: QCA Workshop

Dataset

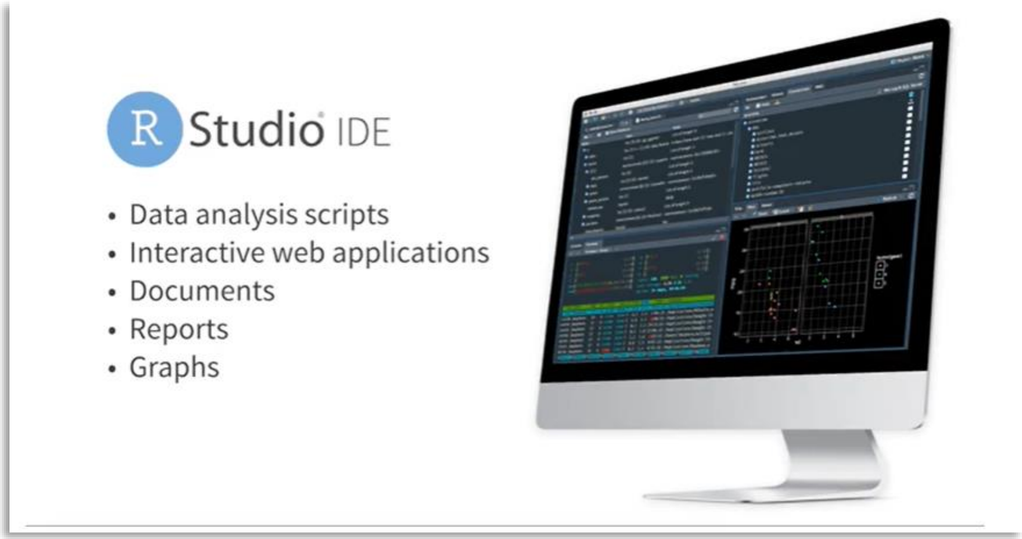
- Survival of democracy (Lipset)
- Built-in QCA package
- Cases are countries
- Attributes are measures of development

Calibrate and analyse

- Crisp set (Boolean) QCA
- Fuzzy set QCA

R packages (+ tidyverse)

- QCA



Alternative software: fsQCA

fsQCA
http://www.socsci.uci.edu/~cragin/fsQCA/software.shtml

Edit Truth Table

msg_qual_c	ben_prsnl_c	prsnl_qual_c	weak_neg_c	weak_pos_c	strong_neg_c	strong_pos_c	number	int_purchase_c	raw consist. ▼	PRI consist.
1	1	1	0	1	0	0	4	1	0.970427	0.909128
0	1	1	0	1	0	1	8	1	0.968773	0.887861
1	1	1	0	1	0	1	65	1	0.965255	0.934447
1	1	1	1	1	1	1	15	1	0.959229	0.876914
1	1	1	0	0	0	1	28	1	0.958419	0.901761
1	1	1	0	0	1	1	4	1	0.957198	0.818826
0	1	1	0	0	0	1				
1	1	0	0	1	0	1				
0	0	1	0	1	0	1				
0	0	1	0	0	0	1				
0	1	1	0	0	1	0				
1	1	1	0	0	0	0				
1	1	0	0	0	0	0				
1	0	1	0	0	0	0				
0	1	1	0	0	0	0				
0	1	0	0	0	0	0				
0	0	1	0	0	1	0				
0	0	0	0	1	0	0				
1	0	0	0	0	0	0				
0	0	0	0	1	0	1				
0	0	1	0	0	0	0				
0	0	0	0	0	1	0				
0	0	0	0	0	0	0				

Reset | Cancel | Specify Analysi

```
**********************
*TRUTH TABLE ANALYSIS*
**********************

File:
Model: int_purchase_c = f(strong_pos_c, strong_neg_c, weak_pos_c, weak_neg_c, prsnl_qual_c, ben_prsnl_c, msg_qual_c)

 Rows:      22


 Algorithm: Quine-McCluskey
      True: 1
  0 Matrix: 0L
Don't Care: -

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 3.000000
consistency cutoff: 0.863937
Assumptions:

                                                                                       raw       unique
                                                                                     coverage   coverage   consistency
                                                                                    ---------- ---------- ----------
~strong_pos_c*~strong_neg_c*~weak_pos_c*~weak_neg_c*ben_prsnl_c                      0.535961   0.051378   0.837382
strong_pos_c*~strong_neg_c*~weak_neg_c*prsnl_qual_c*~msg_qual_c                      0.261382   0.018454   0.932077
~strong_neg_c*~weak_neg_c*prsnl_qual_c*ben_prsnl_c*msg_qual_c                        0.690433   0.052226   0.917665
~strong_pos_c*strong_neg_c*~weak_pos_c*~weak_neg_c*prsnl_qual_c*~msg_qual_c          0.161339   0.007461   0.877363
~strong_pos_c*~strong_neg_c*~weak_pos_c*~weak_neg_c*prsnl_qual_c*msg_qual_c          0.471245   0.002911   0.895735
strong_pos_c*~strong_neg_c*weak_pos_c*~weak_neg_c*ben_prsnl_c*msg_qual_c             0.337968   0.007178   0.950712
strong_pos_c*~weak_pos_c*~weak_neg_c*prsnl_qual_c*ben_prsnl_c*msg_qual_c             0.337572   0.005172   0.956365
~strong_pos_c*~strong_neg_c*weak_pos_c*~weak_neg_c*~prsnl_qual_c*~ben_prsnl_c*~msg_qual_c  0.118073   0.004663   0.863937
strong_pos_c*strong_neg_c*weak_pos_c*weak_neg_c*prsnl_qual_c*ben_prsnl_c*msg_qual_c  0.133644   0.023597   0.959229
solution coverage: 0.840553
solution consistency: 0.840435
```