Qualitative Comparative Analysis

Claude Rubinson
University of Houston—Downtown
Houston, TX
rubinsonc@uhd.edu
http://gator.uhd.edu/~rubinsonc/

76th Annual Meeting of the
Southern Sociological Society
Atlanta, Georgia
April 25, 2013

Overview

- Review of QCA resources, publications, and software
- QCA as an investigation of invariance
- Three analytic components of QCA
  - Data set calibration

    Break
  - Necessity analysis
    - Consistency and coverage measures for necessity
    - Testing for necessary conditions
  - Sufficiency analysis
    - Consistency and coverage measures for sufficiency
    - Constructing and reducing truth tables
    - Interrogating solutions
- Software demonstration

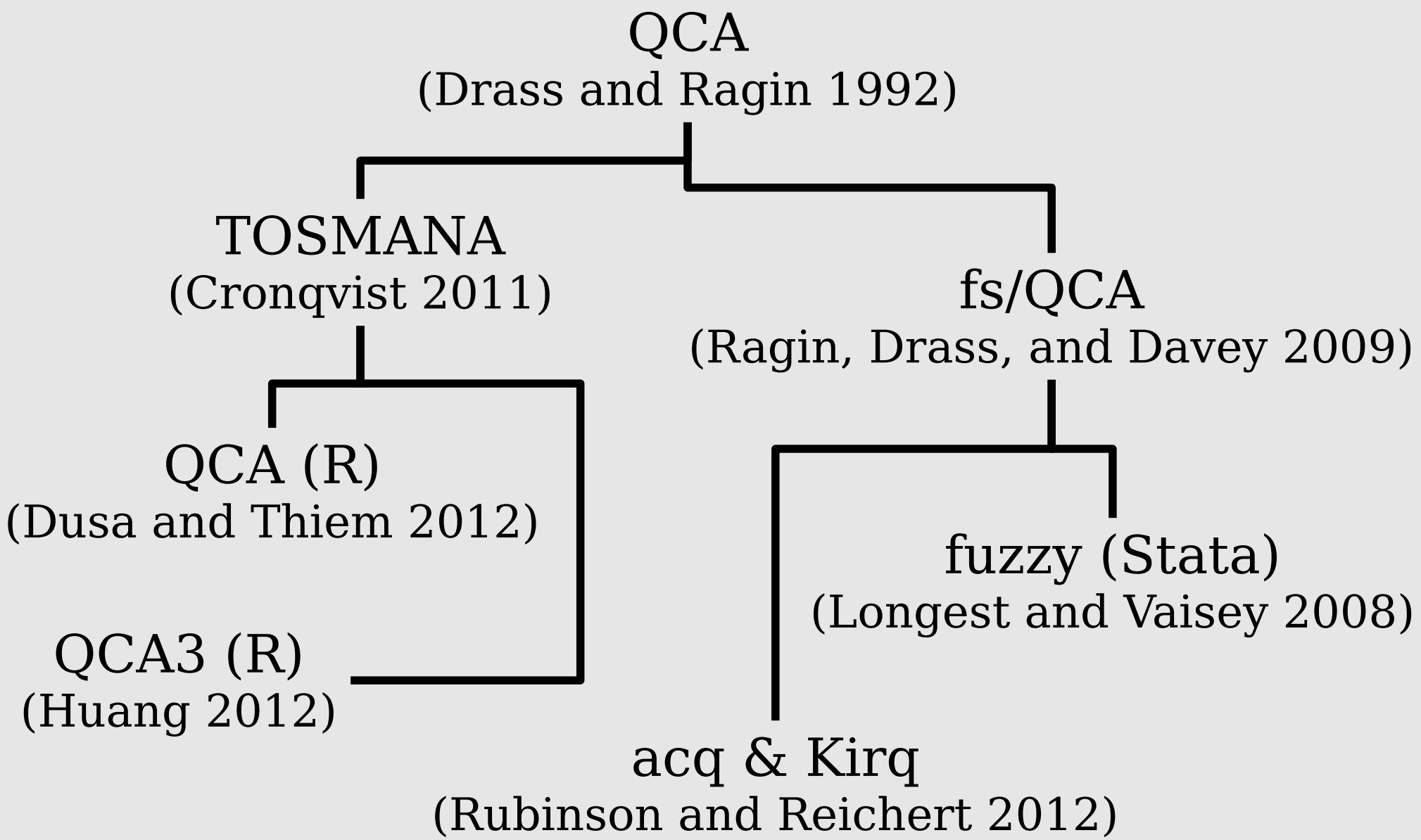
Primary Readings on QCA

- Ragin and Rubinson (2009) “The Distinctiveness of Comparative Research”
- Ragin and Rubinson (2011) “Comparative Methods”
- Ragin (2008) *Redesigning Social Inquiry*
- Ragin (1987) *The Comparative Method*

Secondary Sources

- Compasss web site (http://www.compasss.org)
- Schneider and Wagemann (2012) *Set-Theoretic Methods for the Social Sciences*
- Goertz (2006) *Social Science Concepts*
- Ragin (2000) *Fuzzy-Set Social Science*
- Rihoux and Ragin (2009) *Configurational Comparative Methods*

Software


QCA
(Drass and Ragin 1992)
TOSMANA
(Cronqvist 2011)
fs/QCA
(Ragin, Drass, and Davey 2009)
QCA (R)
(Dusa and Thiem 2012)
QCA3 (R)
(Huang 2012)
fuzzy (Stata)
(Longest and Vaisey 2008)
acq & Kirq
(Rubinson and Reichert 2012)

Software

Ragin's fs/QCA (http://www.fsqca.com):

- always produces correct results, intermediate solutions, relatively user-friendly, described in *RSI,* cutting edge
- but: Windows-only, buggy, out of date documentation, the dreaded prime implicant chart, no tools for interrogating the analysis, cutting edge

Rubinson's acq & Kirq (http://www.grundrisse.org/qca/):

- always produces correct results, sophisticated necessity analysis, supports contradictions and impossible conditions, user-friendly, cross-platform, tools for interrogating the analysis, no prime implicant chart
- but: no intermediate solutions, solution complexity

Software

Cronqvist's TOSMANA:

- visualizations; cross-platform (via Mono)
- but: doesn't support fuzzy-set QCA; inspired by QCA 3.0

Dusa's QCA for R (also Huang's QCA3 for R):

- cross-platform (via R)
- but: no GUI interface; inspired by TOSMANA/QCA 3.0

Longest and Vaisey's fuzzy module for Stata:

- cross-platform (via Stata); focus on probabilistic methods

Varieties of QCA: Crisp Sets, Fuzzy Sets, & Multi-Valued Sets

- *The Comparative Method* (1987) describes "crisp-set QCA"
- *Fuzzy-Set Social Science* (2000) describes "fuzzy-set analysis"
- *Redesigning Social Inquiry* (2008) unifies "crisp-set QCA" and "fuzzy-set QCA"
  - crisp-set QCA is a special form of fuzzy-set QCA
  - fs/QCA, acq, and Kirq are all based on the *RSI* algorithms
- What about multi-valued QCA?

What is QCA?

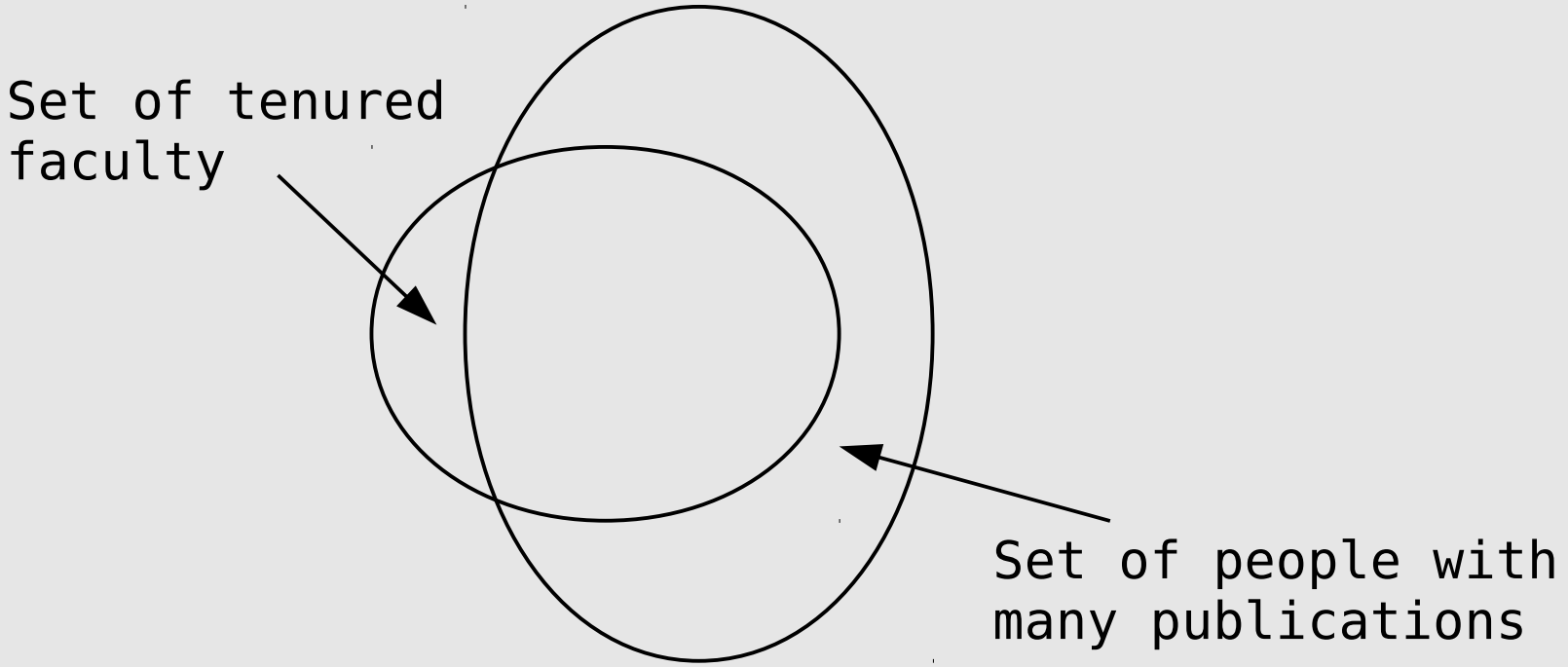
- Originated as a formalization of small-N, comparative-historical research.
- QCA is a technique for identifying and analyzing invariant (consistent) relationships.
- Characterized by the search for necessary and sufficient conditions.
- Is QCA necessarily small-N?
- Is QCA necessarily case-oriented?

Invariant Relationships

- Definition: Certain aspects of cases tend to co-occur.
  - Tenured faculty tend to have many publications
  - Religious fundamentalists tend to be politically conservative
  - “business leaders and owners of capital … are overwhelmingly Protestant” (Weber 1958:35)

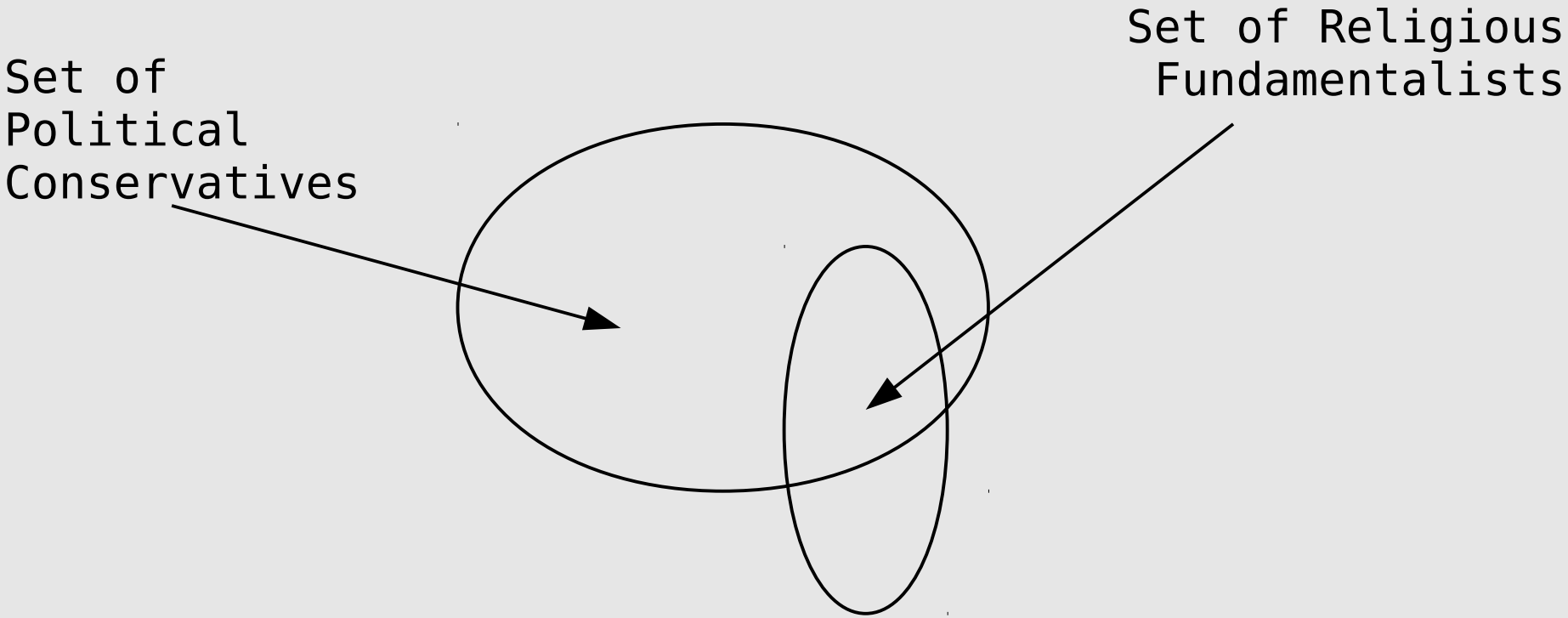
Invariant Relationships

- Definition: Certain aspects of cases tend to co-occur.
  - Tenured faculty tend to have many publications



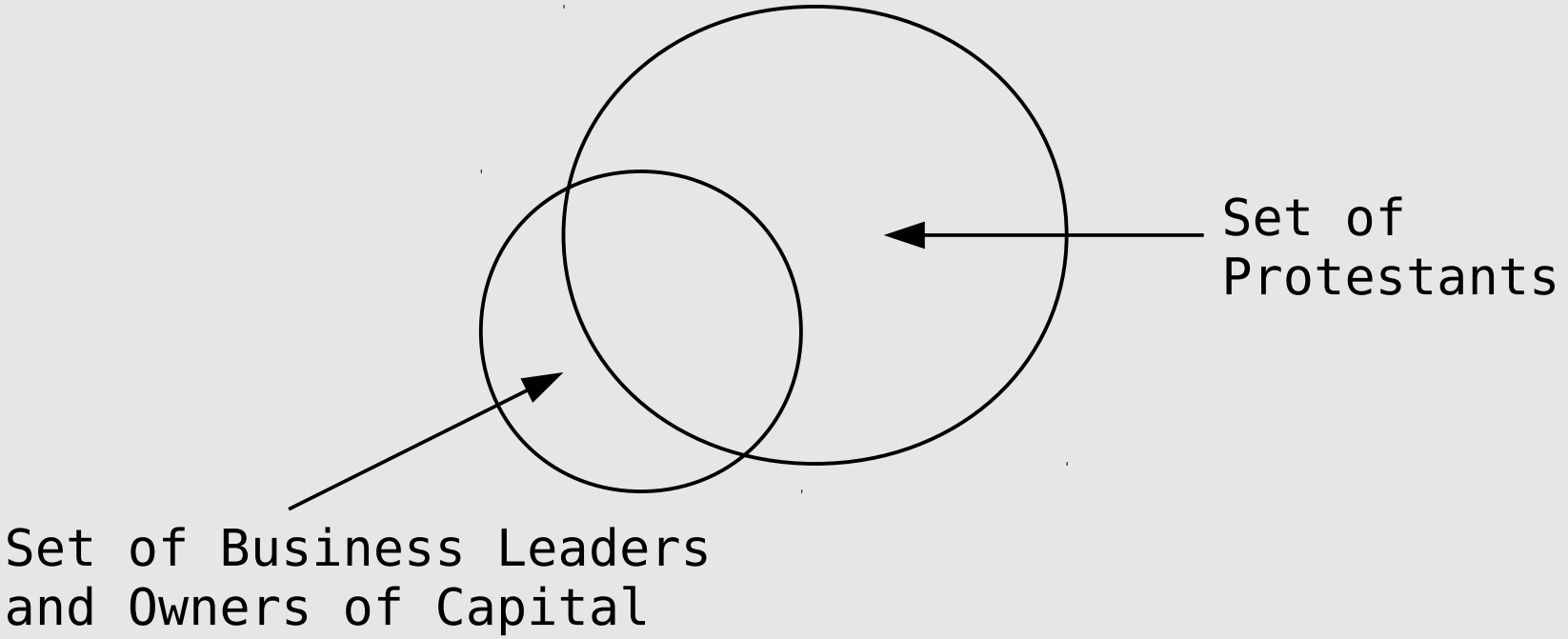
Invariant Relationships

- Definition: Certain aspects of cases tend to co-occur.
    - Religious fundamentalists tend to be politically conservative.

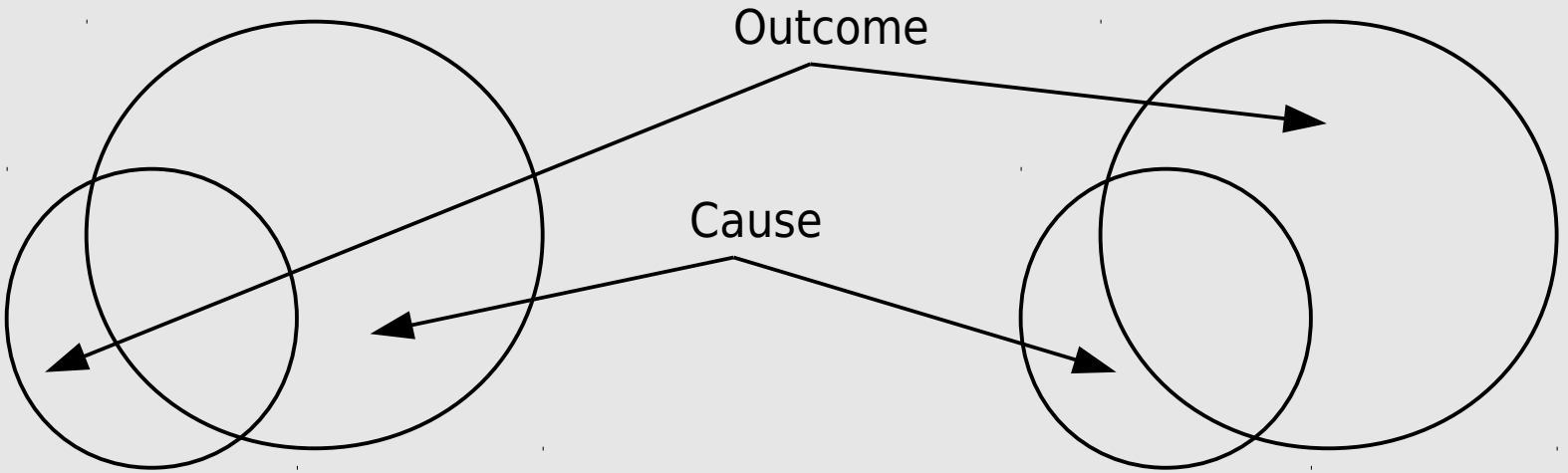


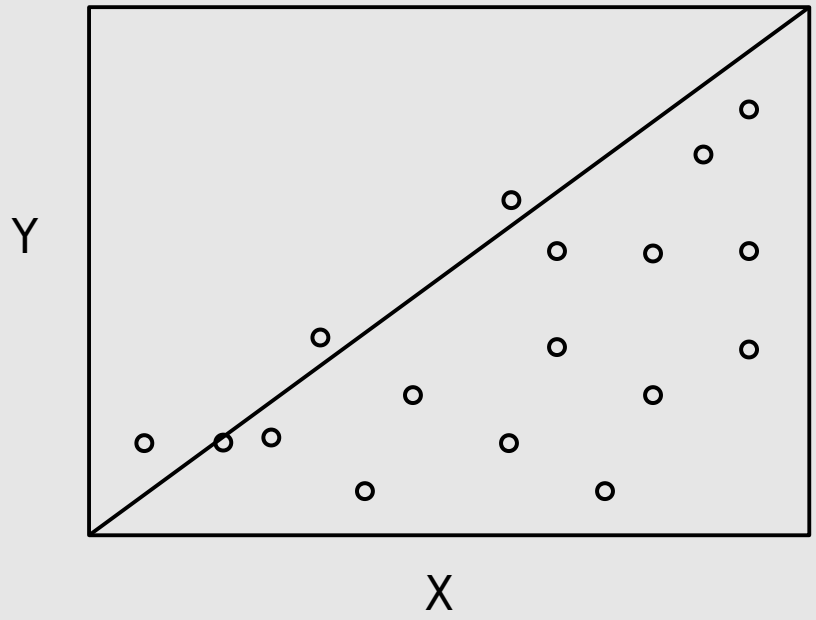
Invariant Relationships

- Definition: Certain aspects of cases tend to co-occur.
  - "business leaders and owners of capital … are overwhelmingly Protestant" (Weber 1958:35)

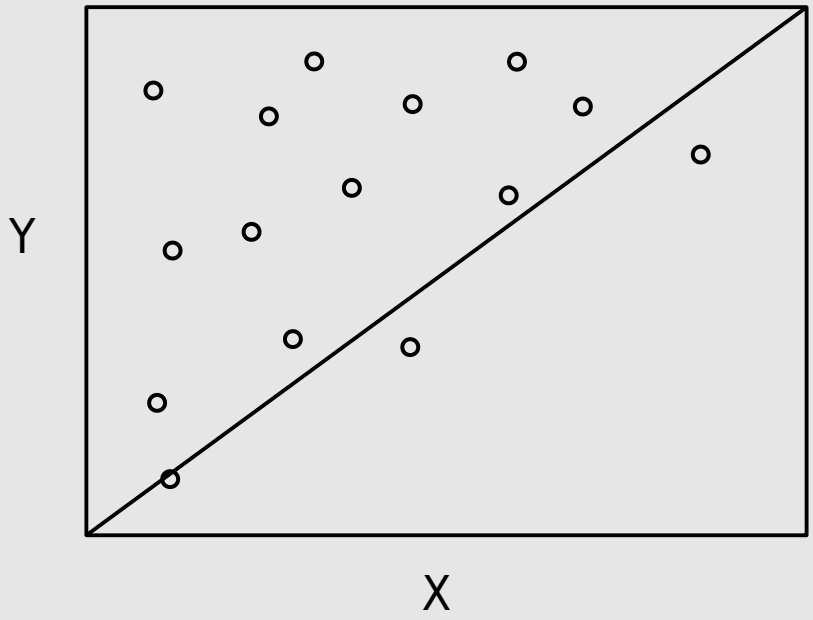


Invariant Relationships





Subset relationship consistent with *necessity* (X≥Y)

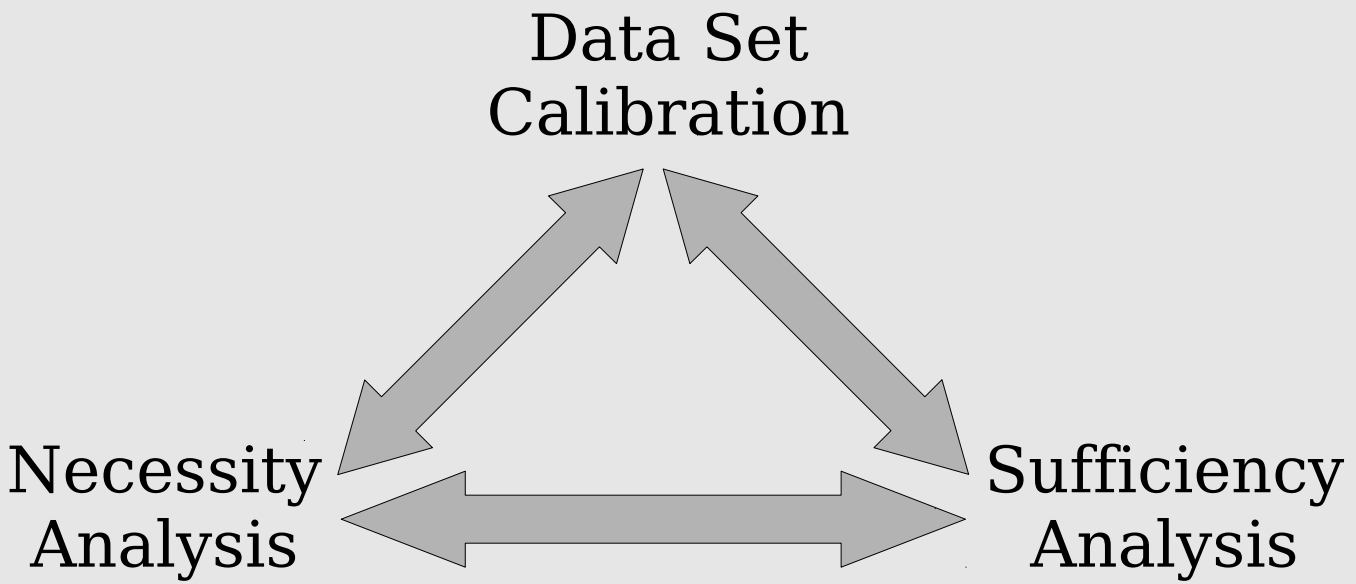


Subset relationship consistent with *sufficiency* (Y≥X)

Invariant Relationships

- Definition: Certain aspects of cases tend to co-occur.
  - Does not imply determinism (or stochasticism)
  - Is not vulnerable to a single disconfirming case
  - Is fundamentally set theoretic
  - Parallels how we typically formulate social theory:
    - The modern world-system is a capitalist world-economy characterized by a single division of labor that prioritizes the endless accumulation of capital.
    - During unsettled periods, people actively use culture to learn new ways of being.

Three Analytic Components of QCA



Retroductive Nature of QCA
Example: Brown and Boswell (1995)

Truth Table with Contradiction (from Table 4 of Brown and Boswell 1995)

Recent Black Migrants	Weak Union	Black Strikebreaking	Observations
T	T	T	East Chicago, Pittsburgh, Youngstown
T	F	Con	Buffalo, Chicago, Gary, Johnstown, [Cleveland]
F	T	F	Bethlehem, Joliet, McKeesport, Milwaukee, New Castle, Reading
F	F	F	Decatur, Wheeling

Retroductive Nature of QCA
Example: Brown and Boswell (1995)

Revised Truth Table without Contradiction (from Table 5 of Brown and Boswell 1995)

Recent Black Migrants	Weak Union	Local Govt Repression	Black Strikebreaking	Observations
T	T	T	T	East Chicago, Pittsburgh, Youngstown
T	T	F	—	
T	F	T	T	Buffalo, Chicago, Gary, Johnstown
T	F	F	F	Cleveland
F	T	T	F	Bethlehem, Joliet, McKeesport, New Castle, Reading
F	T	F	F	Milwaukee
F	F	T	F	Decatur
F	F	F	F	Wheeling

Boolean Algebra

- UPPERCASE for the presence of a condition
- lowercase for the absence of a condition
- Negation

  ~A = 1 – A

  a = 1 – A

- Logical and (Boolean multiplication)

  A•b = Ab = min(A,b)

- Logical or (Boolean addition)

  A+b = max(A,b)

Calibrating Data Sets

Data Set Calibration

- The process of constructing fuzzy-sets
- May be crisp or fuzzy
- Is about defining set memberships
  - degree of membership in the set of rich people (vs annual income)
  - degree of membership in the set of developed countries (vs GDP/capita)
- Importance of negation and asymmetry
  - degree of membership in the set of *not* rich people
  - degree of membership in the set of *not* developed countries

Data Set Calibration

- Instrument calibration is routine in the natural sciences; largely absent in the social sciences.
- Conventional statistics emphasize relative effects: Paul is poorer than Peter; the United States' infant mortality rate is greater than that of Japan.
- Calibration allows us to state that an individual is poor or that a country's infant mortality rate is high.
- Calibration requires application of theoretical and substantive knowledge: What does it mean to classify a country as partially versus fully democratic?

Calibrating Fuzzy Sets

Crisp set	Three-value fuzzy set	Four-value fuzzy set	Six-value fuzzy set	Continuous fuzzy set
1 = fully in	1 = fully in	1 = fully in	1 = fully in	1 = fully in
	0.67 = more in than out	0.67 = more in than out	0.8 = mostly but not fully in 0.6 = more or less in	Degree of membership is more “in” than “out” 0.5 < X < 1
---------- 0.5 = Crossover Point ----------				
		0.33 = more out than in	0.4 = more or less out 0.2 = mostly but not fully out	Degree of membership is more “out” than “in” 0.0 < X < 0.5
0 = fully out	0 = fully out	0 = fully out	0 = fully out	0 = fully out

Analyzing Necessary Conditions

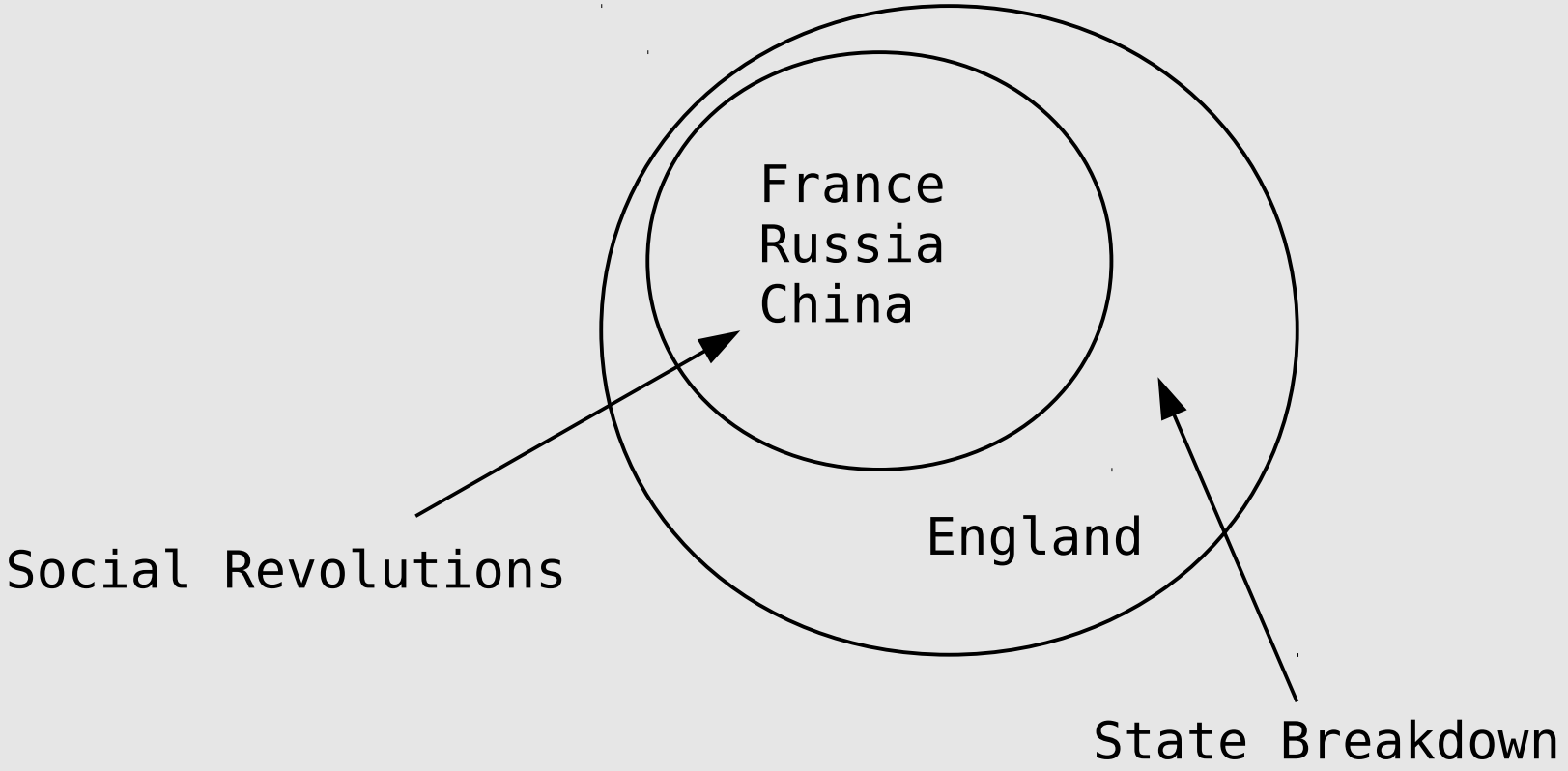
Necessity Analysis

- Underdeveloped in the literature; QCA development has focused on sufficiency analysis
- Kirq and acq have sophisticated necessity testing

Necessary Conditions

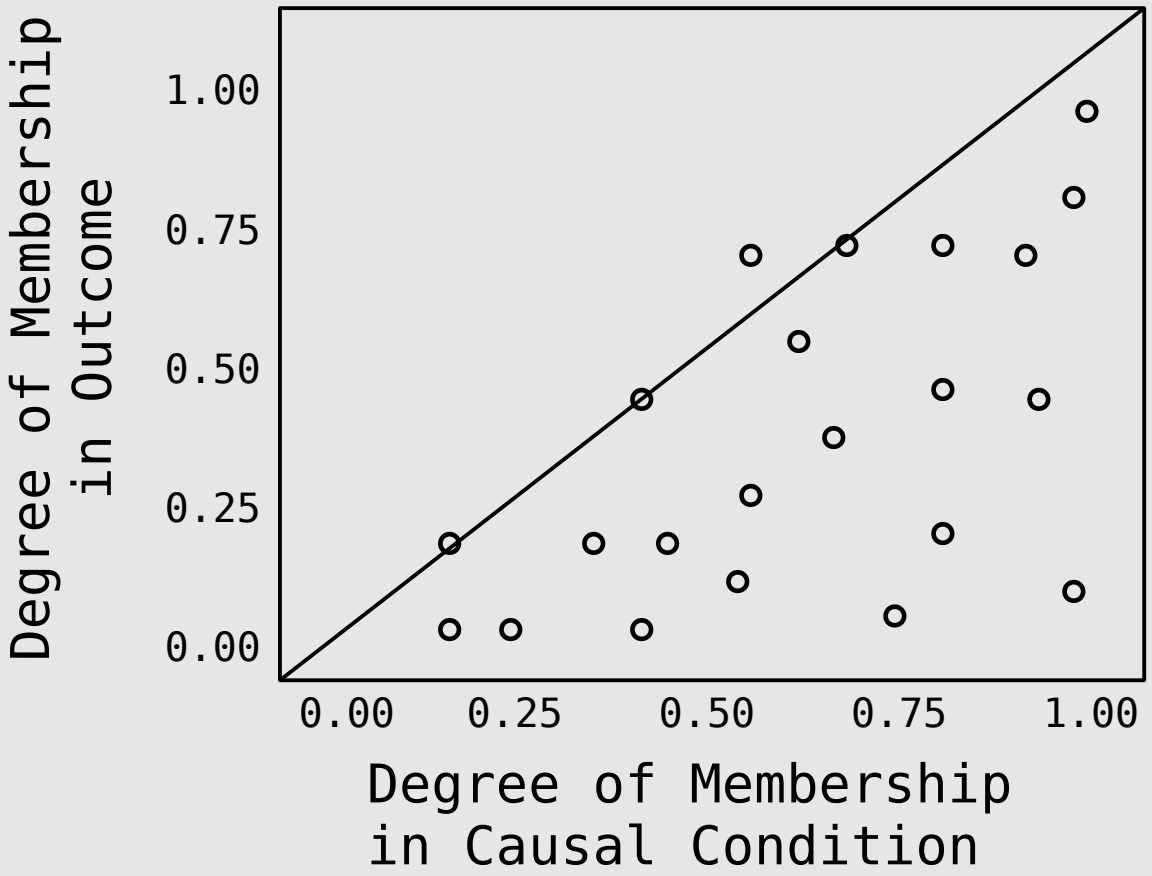
Causal condition must (almost always) be present for outcome to occur.

Outcome is a subset of Cause



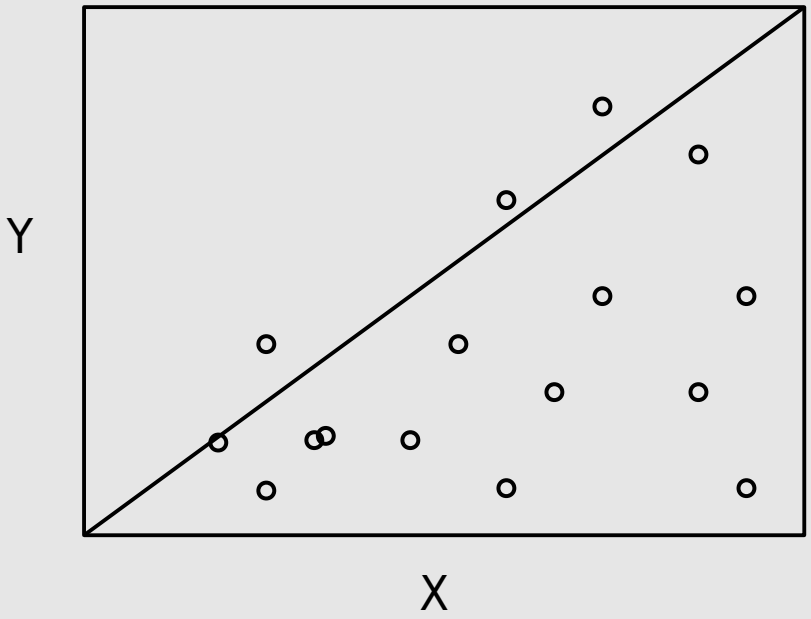
Fuzzy Subset Relationship Consistent with Necessity

Outcome is a subset of Cause ($X \geq Y$)

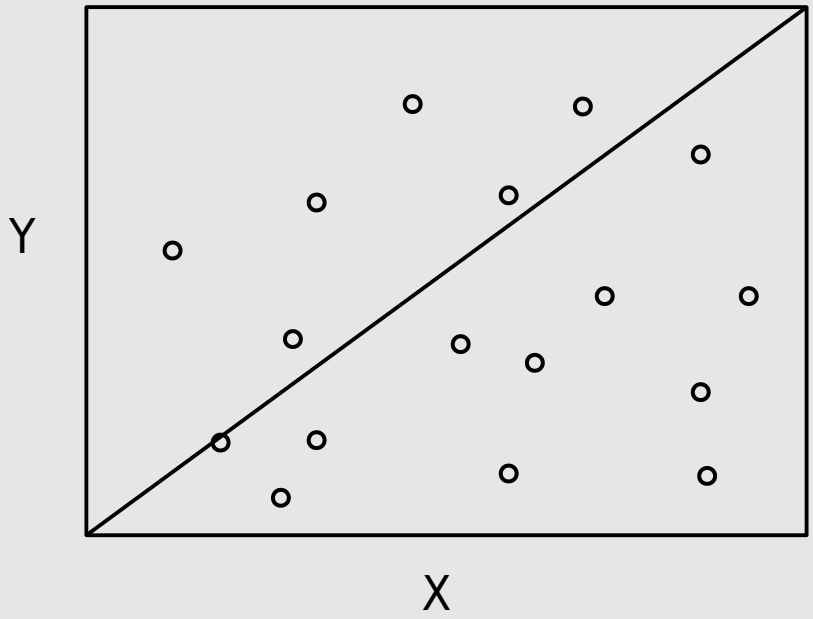


Assessing Necessary Conditions

- *Consistency* measures degree to which subset relationship is “consistent” with necessity



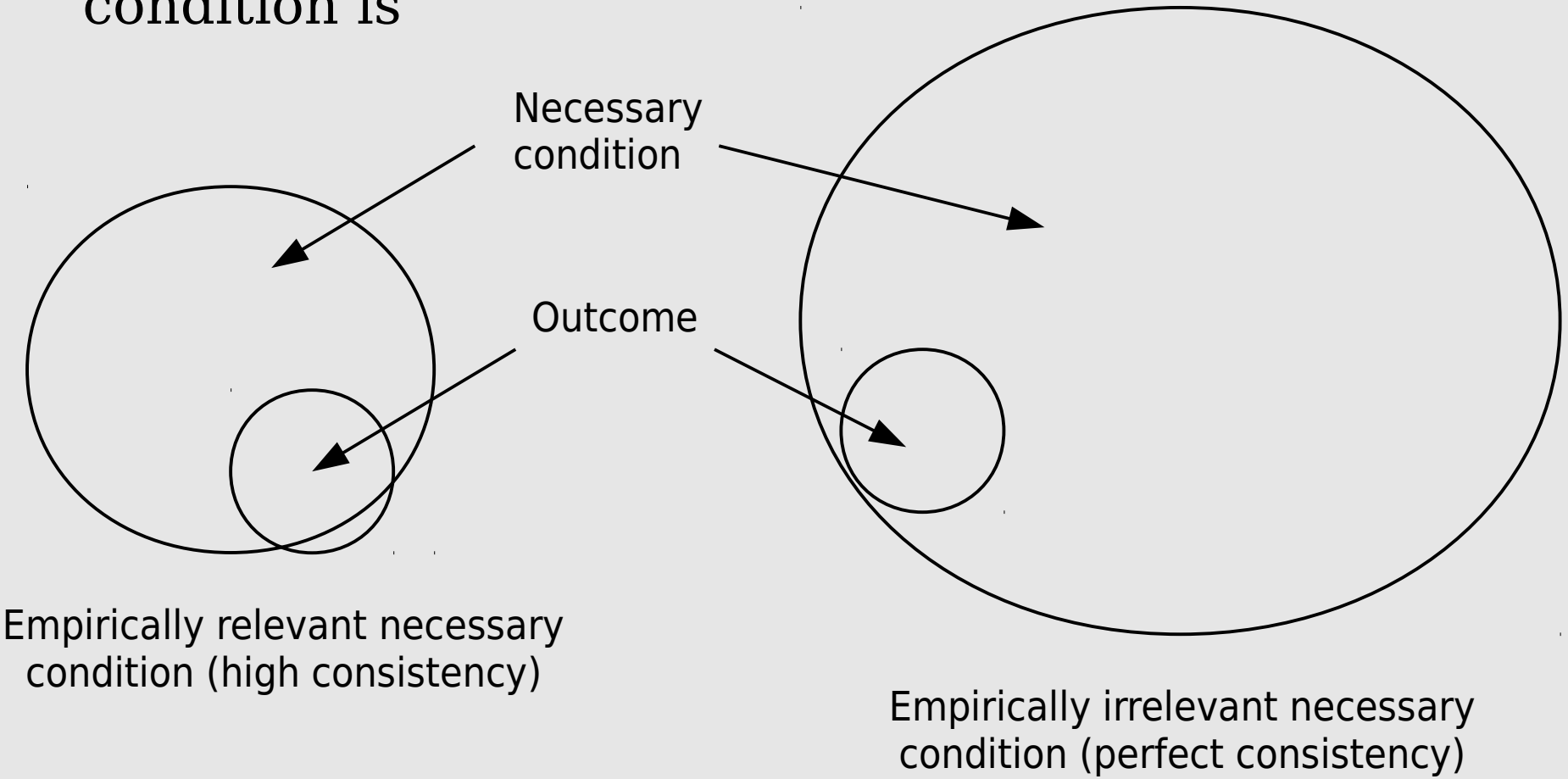
Subset relationship consistent with necessity



Subset relationship with substantial inconsistency

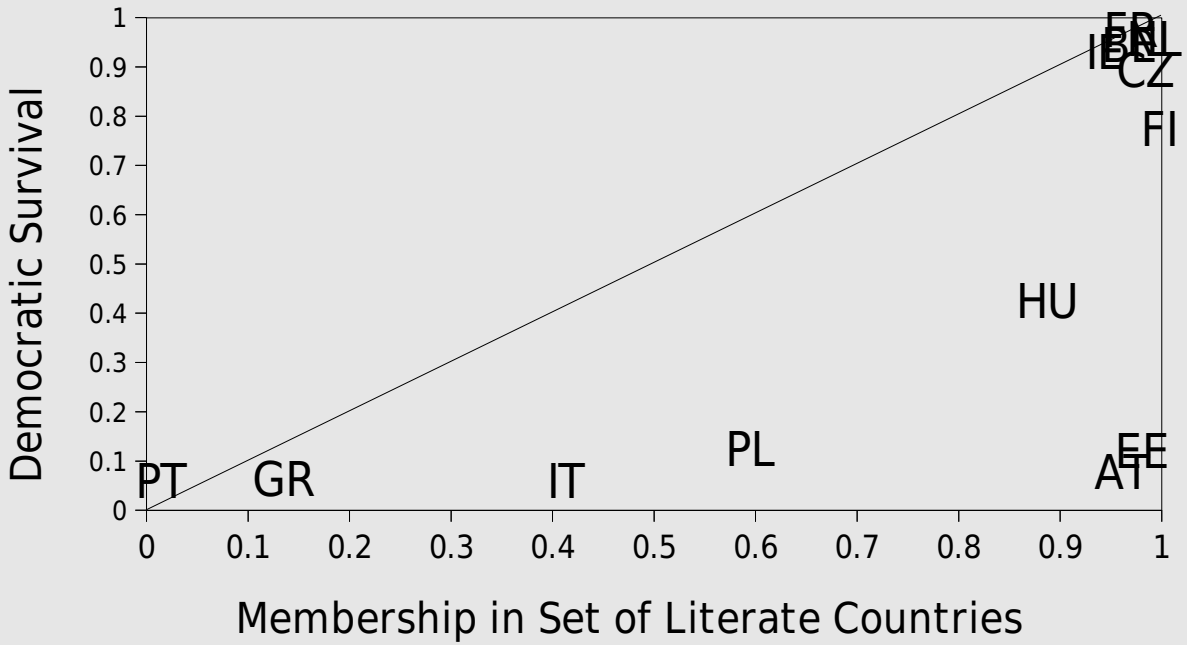
Assessing Necessary Conditions

- *Coverage* measures how "relevant" a necessary condition is



Testing for Necessary Conditions

Obs	Dev	Urb	Lit	Sur
AT	.81	.12	.99	.05
BE	.99	.89	.98	.95
CZ	.58	.98	.98	.89
EE	.16	.07	.98	.12
FI	.58	.03	.99	.77
FR	.98	.03	.99	.95
DE	.89	.79	.99	.05
GR	.04	.09	.13	.06
HU	.07	.16	.88	.42
IE	.72	.05	.98	.92
IT	.34	.10	.41	.05
NL	.98	1.00	.99	.95
PL	.02	.17	.59	.12
PT	.01	.02	.01	.05



Term	Consis	Cov
LIT	0.99	0.58
Solution	0.99	0.58

Testing for Necessary Conditions

- Assess consistency before coverage
- Many solutions are possible
- Use of theory and substantive knowledge is crucial

Analyzing Sufficient Conditions

Sufficiency Analysis

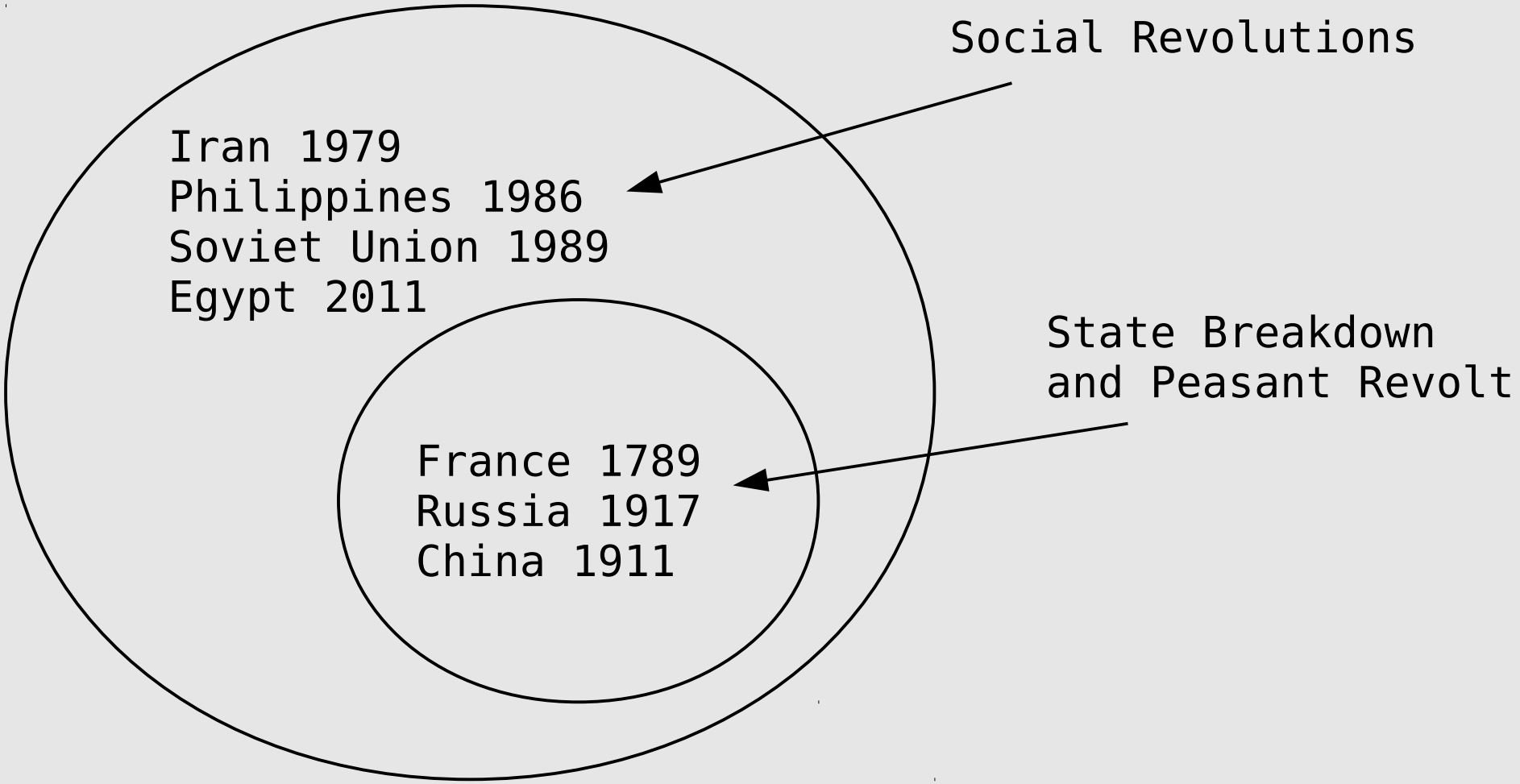
- More mature than necessity analysis; QCA development—and applications—have focused on sufficiency analysis
- Emphasis on causal complexity (a.k.a., multiple conjunctural causation, “recipes,” equifinality, or INUS conditions)

Feature	fs/QCA	Kirq & acq
Based on RSI Algorithms	√	√
Complex Solutions	√	√
Intermediate Solutions	√	
Parsimonious Solutions	√	√
Impossible Conditions		√
Contradictions		√

Sufficient Conditions

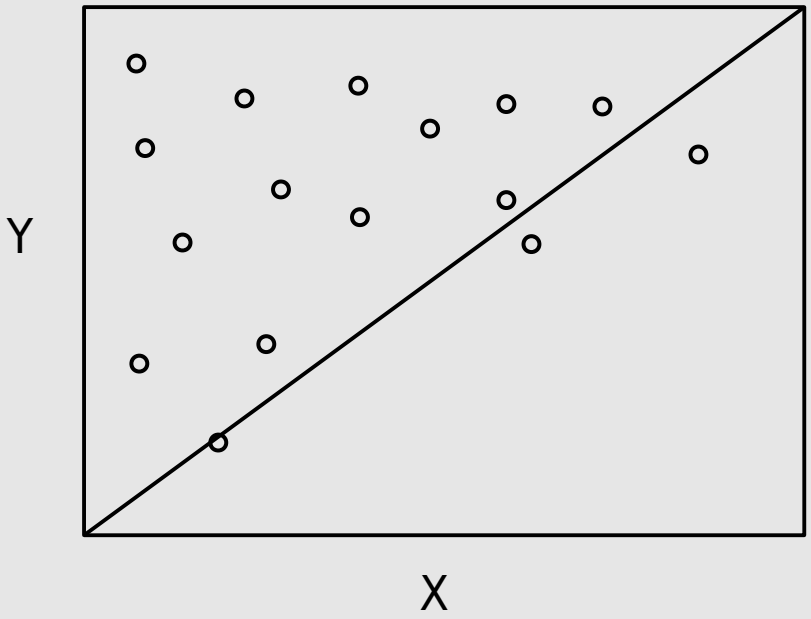
Outcome (almost) always occurs when causal condition is present.

Cause is a subset of Outcome

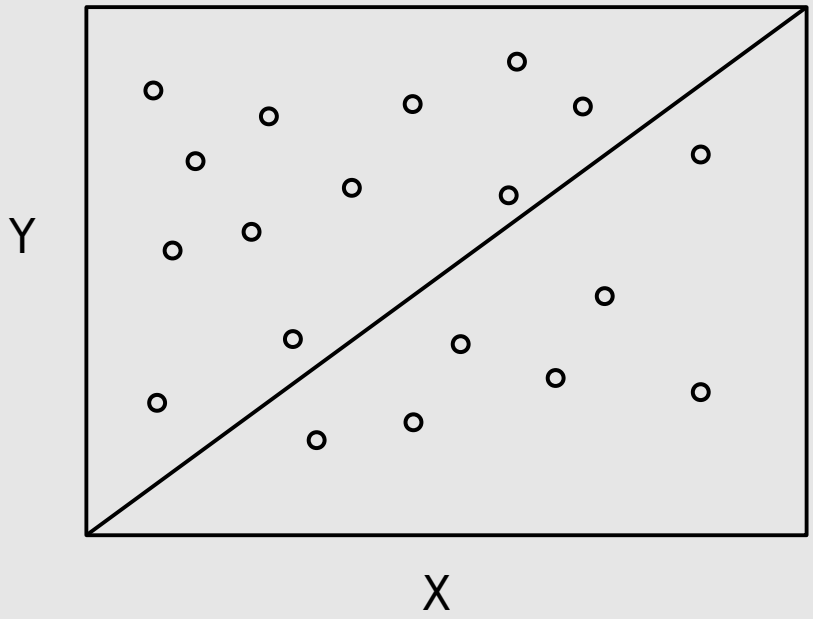


Assessing Sufficient Conditions

- *Consistency* measures degree to which subset relationship is “consistent” with sufficiency



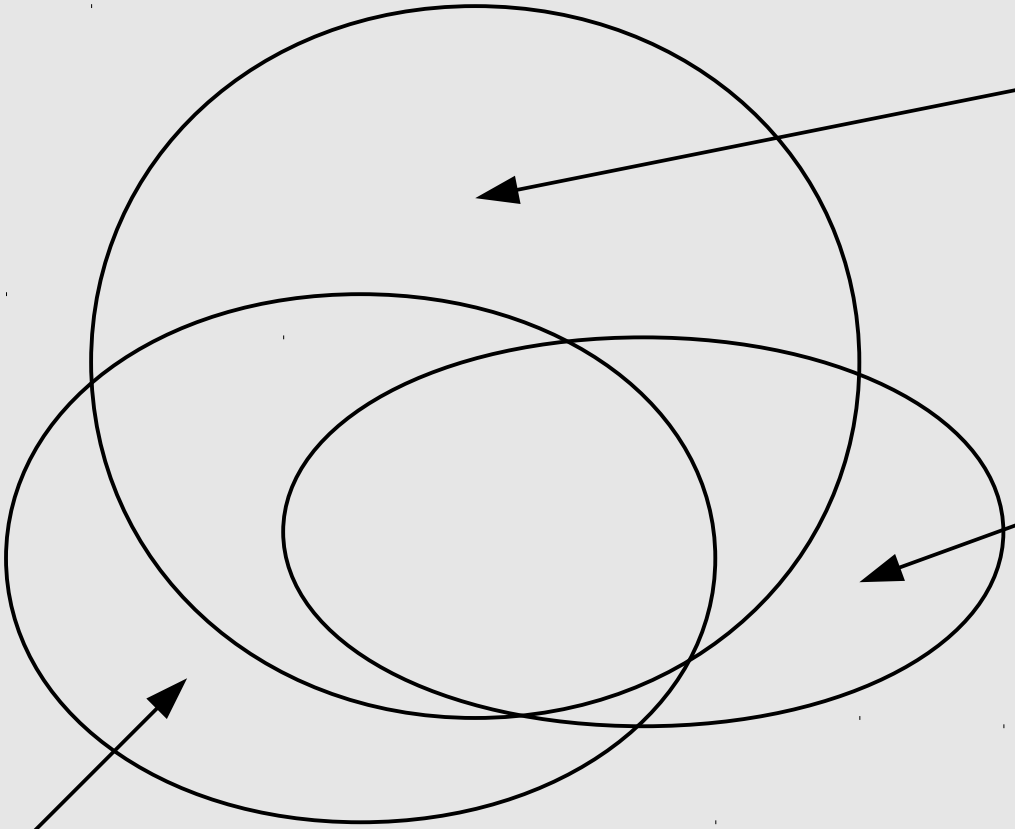
Subset relationship consistent with sufficiency



Subset relationship with substantial inconsistency

Assessing Sufficient Conditions

- *Coverage* measures the relative "importance" of each solution



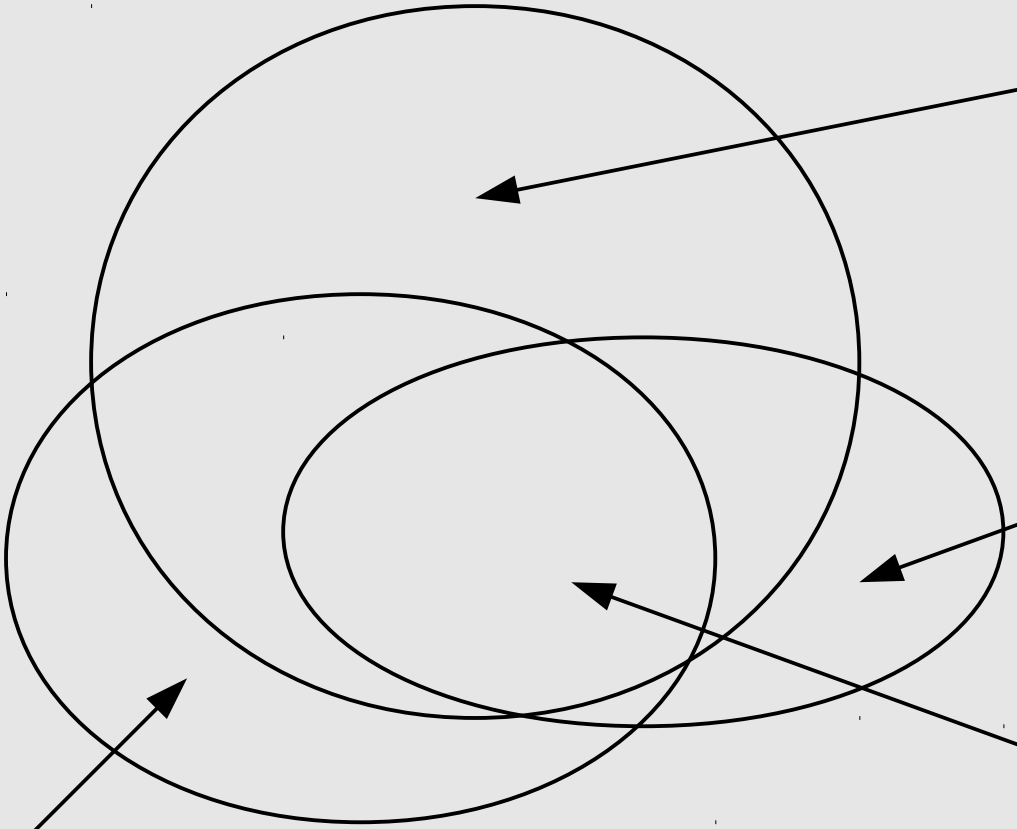
Recently deported women who do not plan to cross again (Outcome)

High SES women who haven't lived in the U.S. and aren't traveling with family (X_1)

High SES women who haven't lived in the U.S., have only attempted cross a few times and felt that their last crossing experience was very dangerous (X_2)

Assessing Sufficient Conditions

- *Coverage* measures the relative "importance" of each solution



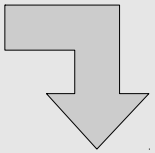
Testing for Sufficient Conditions

Term	Consis	Raw Cov	Uniq Cov
HISES*liveus*travfam +	0.90	0.32	0.13
HISES*liveus*numcross*DANGER	0.82	0.48	0.26
Solution	0.86	0.58	

Truth Table Construction

Truth table algorithm sorts observations into types

Obs	Dev	Urb	Lit	Brk
AT	.81	.12	.99	.95
BE	.99	.89	.98	.05
CZ	.58	.98	.98	.11
EE	.16	.07	.98	.88
FI	.58	.03	.99	.23
FR	.98	.03	.99	.05
DE	.89	.79	.99	.95
GR	.04	.09	.13	.94
HU	.07	.16	.88	.58
IE	.72	.05	.98	.08
IT	.34	.10	.41	.95
NL	.98	1.00	.99	.05
PL	.02	.17	.59	.88
PT	.01	.02	.01	.95



	Dev	Urb	Lit	Consis	Y	Consis Obs	Inconsis Obs
1	T	T	T	0.41	F	DE	BE, CZ, NL
2	T	T	F	—	—		
3	T	F	T	0.51	F	AT	FI, FR, IE
4	T	F	F	—	—		
5	F	T	T	—	—		
6	F	T	F	—	—		
7	F	F	T	0.83	T	EE, PL	HU
8	F	F	F	0.99	T	GR, IT, PT	

Reading Truth Tables

Truth table assesses consistency between types and outcome

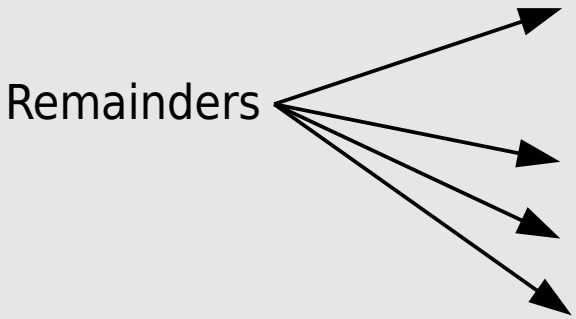
Democracy usually did not break down in countries that were (a) developed, urbanized, and literate (row 1) or (b) developed, not urbanized, and literate (row 3).

Democracy usually did break down in countries that were (c) not developed, not urbanized, and literate (row 7) or (d) not developed, not urbanized, and not literate (row 8)

	Dev	Urb	Lit	Consis	Y	Consis Obs	Inconsis Obs
1	T	T	T	0.41	F	DE	BE, CZ, NL
2	T	T	F	—	—		
3	T	F	T	0.51	F	AT	FI, FR, IE
4	T	F	F	—	—		
5	F	T	T	—	—		
6	F	T	F	—	—		
7	F	F	T	0.83	T	EE, PL	HU
8	F	F	F	0.99	T	GR, IT, PT	

Reading Truth Tables

Remainders are logically possible conditions lacking empirical instances



	Dev	Urb	Lit	Consis	Y	Consis Obs	Inconsis Obs
1	T	T	T	0.41	F	DE	BE, CZ, NL
2	T	T	F	—	—		
3	T	F	T	0.51	F	AT	FI, FR, IE
4	T	F	F	—	—		
5	F	T	T	—	—		
6	F	T	F	—	—		
7	F	F	T	0.83	T	EE, PL	HU
8	F	F	F	0.99	T	GR, IT, PT	

Invariance in Truth Tables

	Dev	Urb	Consis	Y	Consis Obs	Inconsis Obs
1	T	T	0.41	F	DE	BE, CZ, NL
2	T	F	0.51	F	AT	FI, FR, IE
3	F	T	—	—		
4	F	F	0.89	T	EE, GR, IT, PL, PT	HU

	Dev	Urb	Lit	Consis	Y	Consis Obs	Inconsis Obs
1	T	T	T	0.41	F	DE	BE, CZ, NL
2	T	T	F	—	—		
3	T	F	T	0.51	F	AT	FI, FR, IE
4	T	F	F	—	—		
5	F	T	T	—	—		
6	F	T	F	—	—		
7	F	F	T	0.83	T	EE, PL	HU
8	F	F	F	0.99	T	GR, IT, PT	

Reducing Truth Tables to Boolean Equations

To Primitive Expressions:

Term	Consis	Raw Cov	Uniq Cov	Observations
dev*urb*LIT +	0.83	0.42	0.27	EE, PL, [HU]
dev*urb*lit	0.99	0.40	0.24	GR, IT, PT
Solution	0.88	0.66		

Reducing Truth Tables to Boolean Equations

To Primitive Expressions:

Term	Consis	Raw Cov	Uniq Cov	Observations
dev*urb*LIT +	0.83	0.42	0.27	EE, PL, [HU]
dev*urb*lit	0.99	0.40	0.24	GR, IT, PT
Solution	0.88	0.66		

To Prime Implicants:

Term	Consis	Raw Cov	Uniq Cov	Observations
dev*urb	0.89	0.71	0.71	EE, PL, GR, IT, PT, [HU]
Solution	0.89	0.71		

Reducing Truth Tables to Boolean Equations

Reduce Prime Implicants (Complex Solution):

Term	Consis	Raw Cov	Uniq Cov	Observations
dev*urb	0.89	0.71	0.71	EE, PL, GR, IT, PT, [HU]
Solution	0.89	0.71		

Reducing Truth Tables to Boolean Equations

Reduce Prime Implicants (Complex Solution):

Term	Consis	Raw Cov	Uniq Cov	Observations
dev*urb	0.89	0.71	0.71	EE, PL, GR, IT, PT, [HU]
Solution	0.89	0.71		

Reduce Prime Implicants Using Remainders (Parsimonious Solution):

Term	Consis	Raw Cov	Uniq Cov	Observations
dev	0.82	0.73	0.73	EE, PL, GR, IT, PT, [HU]
Solution	0.82	0.73		

Constructing Intermediate Solutions

Complex Solution

Acsir +
ACSir +
ASIR

Parsimonious Solution

i +
SR

Intermediate Solution #1

Ai +
ACSi +
ASR

Ai +
ASR

Intermediate Solution #2

Air +
ASIR

Factoring Results

Initial Solution:

ELECTIONS * POLICE +
urban * POLICE +
CONFLICT * ELECTIONS * URBAN +
CONFLICT * elections * urban +
conflict * ELECTIONS * urban

Factored Solution:

POLICE (ELECTIONS + urban) +
URBAN (CONFLICT * ELECTIONS) +
urban ((CONFLICT * elections) + (conflict * ELECTIONS))