

Theoretical Foundations of Set Theoretic Methods (STM) and Qualitative Comparative Analysis (QCA)

Rodney Lacey
University of California, Davis

Set Theoretic Methods Professional Development Workshop

Special thanks to Charles Ragin (University of Arizona) and Peer Fiss (University of Southern California) for providing material adapted or borrowed throughout this presentation

Overview

- How do set-theoretic methods (STM) and qualitative comparative analysis (QCA) differ from other methods?
 - Different epistemological and ontological foundations
 - Neither qualitative or quantitative
 - Not just a technique, but rather, a way of theorizing
- What kind of phenomena can you investigate?
- What kind of theories can you build?
- How does the method correspond to the social world?

Foundations

- Comparative analysis of cases
 - In-depth analysis of the directly comparable
 - Theory building and refining simultaneously
- Configurations and set memberships
 - Model the multiple, conjunctive causality of social world
- Diversity
 - Capture and analyze the complexity of social structure as constraints and options

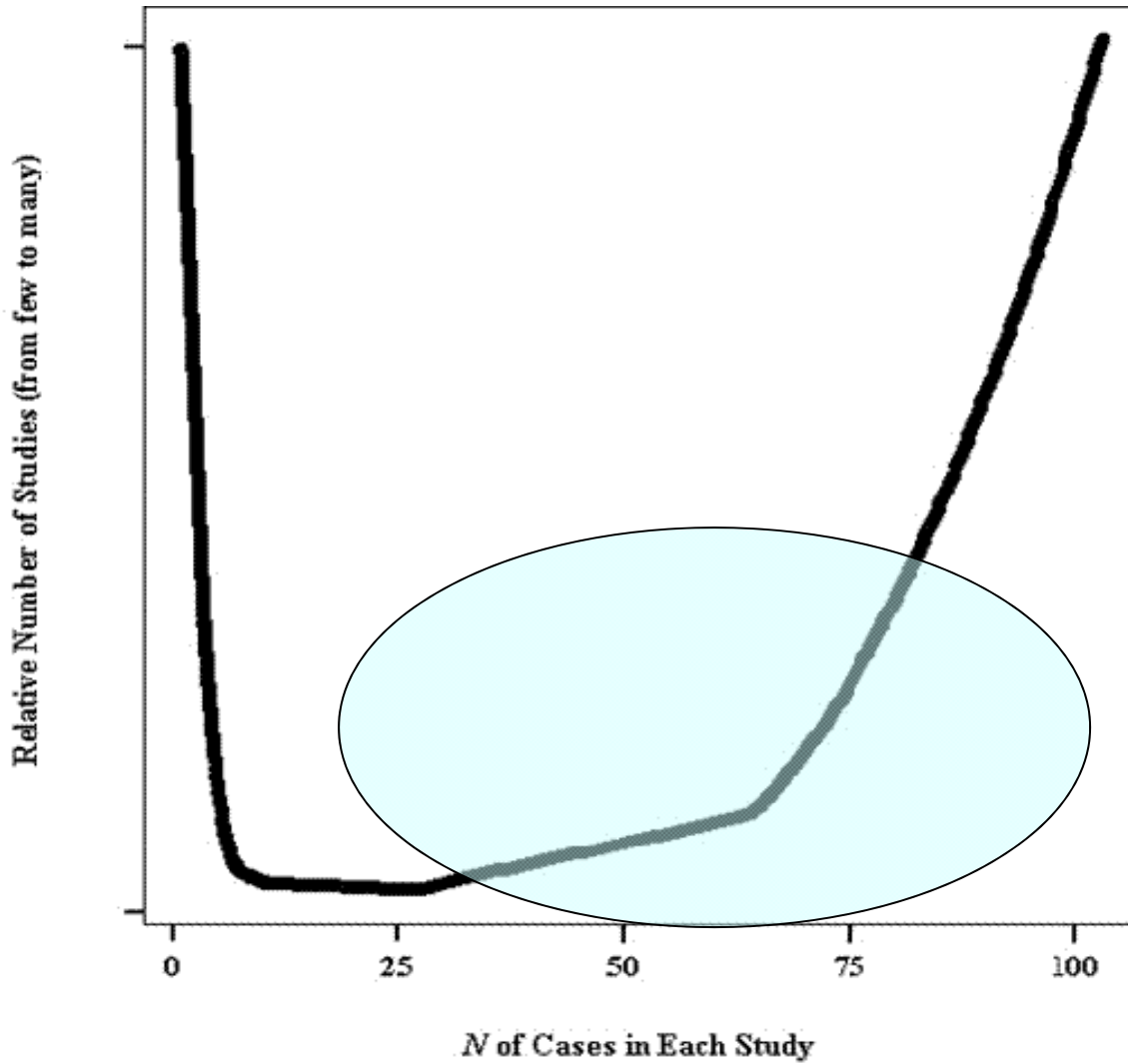
Part I

Comparative analysis of cases

Comparative analysis of cases

- What exactly is the phenomenon?
- Which cases are similar enough to be compared theoretically?
 - Explicitly identify substantively comparable cases
- NOT samples and populations
- In-depth comparisons of comparable cases to build theory

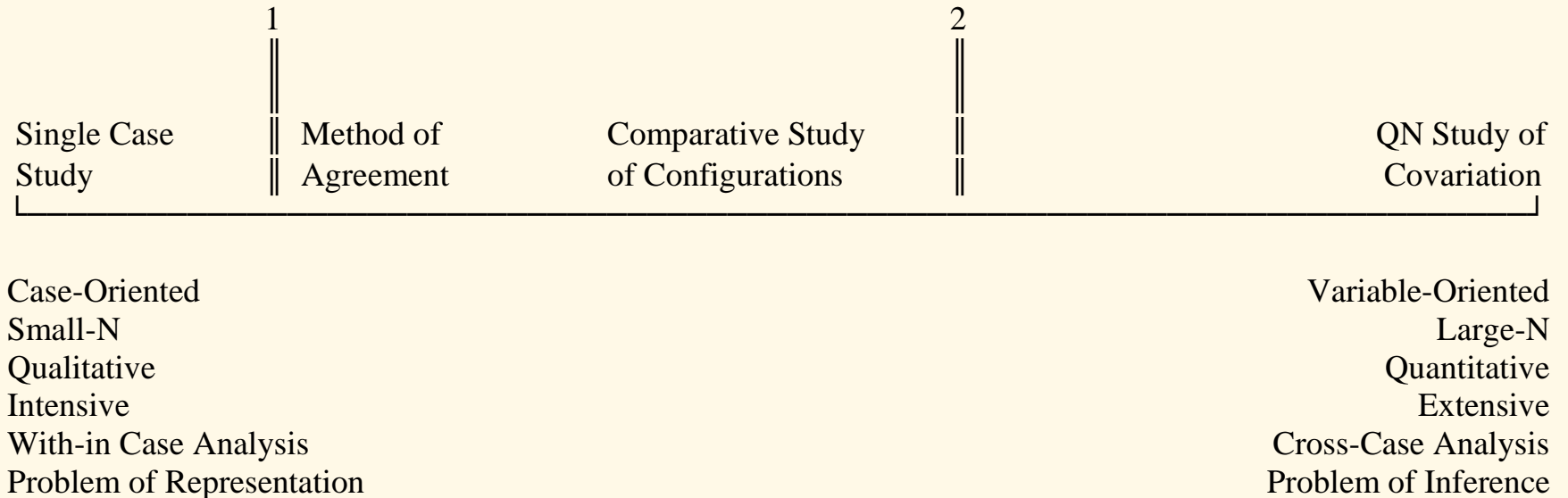
Theorizing the Comparable Case



Source: Ragin (2000:25)

Figure 1.1 Plot of Relative Number of Studies against N of Cases in Each Study

Case-oriented vs. Variable-oriented



Contrasting methods

Quantitative Methods	Comparative Method
General patterns in big populations	Make sense of a few, very similar cases
Population is given, homogeneity assumed, heterogeneity randomly distributed	“What is a case?” is a theoretical discovery, not assumption or test
Larger the population the better	Fewer substantively important cases
Presumed goal is theory-testing	Theory informs analysis reciprocally

Limited view from the regression line

CONVENTIONAL VIEW		
	Cause absent	Cause present
Outcome present	cases in this cell (#1) contribute to error	many cases should be in this cell (#2)
Outcome absent	many cases should be in this cell (#3)	cases in this cell (#4) contribute to error

Necessity and Sufficiency

CAUSE IS NECESSARY BUT NOT SUFFICIENT		
	Cause absent	Cause present
Outcome present	1. no cases here	2. cases here
Outcome absent	3. not relevant	4. not relevant

CAUSE IS SUFFICIENT BUT NOT NECESSARY		
	Cause absent	Cause present
Outcome present	1. not relevant	2. cases here
Outcome absent	3. not relevant	4. no cases here

Necessity (without sufficiency)

I. Expressed as a simple truth table:

<u>Cause</u>		<u>Outcome</u>
1		1
1		0
0		0

II. Expressed as an inequality:

(values of the outcome) \leq (value of the cause)

III. Expressed as a research strategy: Find instances of the outcome (i.e., select on the dependent variable) and assess their agreement on the causal condition (i.e., make sure that the cause does not vary substantially across instances of the outcome).

Sufficiency (without necessity)

I. Expressed as a simple truth table:

<u>Cause</u>		<u>Outcome</u>
1		1
0		1
0		0

II. Expressed as an inequality:

(values of the cause) \leq (value of the outcome)

III. Expressed as a research strategy: Find instances of the causal condition (i.e., select on the independent variable) and assess their agreement on the outcome (i.e., make sure that the outcome does not vary substantially across instances of the cause). This strategy is central to most forms of qualitative research.

Causal model

Example: a researcher studies production sites in a strike-prone industry and considers four possible causes of strikes:

technology = the introduction of new technology

wages = stagnant wages in times of high inflation

overtime = reduction in overtime hours

sourcing = outsourcing portions of production

• is logical “and”

+ is logical “or”

Possible findings include:

(1) *technology* → strikes

(2) *technology* • *wages* → strikes

(3) *technology* + *wages* → strikes

(4) *technology* • *wages* + *overtime* • *sourcing* → strikes

In (1) *technology* is necessary and sufficient

(2) *technology* is necessary but not sufficient

(3) *technology* is sufficient but not necessary

(4) *technology* is neither necessary nor sufficient

Part II

Configurations and set memberships

- Social differences of interest are usually set memberships
- NOT independent effects
- Causality is most often conjunctive and multi-dimensional
- Configurations of set memberships can capture complexity of social causation
 - Necessity and sufficiency
 - Multiple pathways to same outcome

Sets (con't)

- Differences in kind, not amount
 - No marginal effects
- Causality is multiple and conjunctive
 - Holistic combination of set memberships smallest unit of sufficient or necessary cause
 - Single set membership not necessarily independent causal force
- Set membership is **NOT** a limited variance variable
 - Rejects “general linear reality” (Abbott, 1988)

Differences that matter

- Political party, vote, neighborhood, married, children, homeowner, working class, middle class, upper class
- Industry, merger, acquisition, hiring, firing, promoting, growing/shrinking, Fortune 500 and 1000, successful product launch, buy/sell/hold
- Finance CEO, outside/inside CEO or board member, male or female, citizen or non-citizen
-

Money is a set, not a number

- Classically considered to be a continuous variable, money is usually not when it is a social force
- Cultural meanings matter and they are not continuous
- Money generally makes a difference by changing your set membership, not by incremental addition of dollars
 - Higher profits vs. lower profits than last year
 - Grow faster than industry average vs. slower
 - Income \$100 above assistance standard vs. \$1 below
 - Earning \$20K as graduate student vs. \$20k in fast food

Multiple conjunctive causality

“For example, a White, suburban, middle-class, male professional with three children and a mortgage who votes for the Republican Party differs by only a single trait from a person with all these same characteristics but who votes for the Socialist Labor Party. Thus, ... [though] they are very similar ... it might be very different to be trapped in an elevator with one versus the other. This qualitative difference captures the essence of viewing cases as configurations: two cases may be similar in most ways but because they differ on one or more key aspects, their difference may be one of kind, not simply one of degree.” (Ragin, 2000: 71)

- Configurations in organization theory (Fiss, 2007)

Part III

Diversity

Diversity

- Social structure defines options, opportunities
- What is not available often as important as what is
- Rare cases – “outliers” -- matter because they are the boundary of available diversity
- Measured relative to theoretically possible, not empirically observable options
 - Diversity of American presidents low only relative to possible options

Variance is not diversity

- Weighted average, not options
- Outliers are error terms, not part of diversity
- No theoretically derived frame for calculating variance
 - Measure of what is empirically observable
- Clustering techniques using variance reduction are ad hoc, simplistic means of representing diversity
 - How many clusters are enough to adequately capture diversity?

Organizational theories of diversity

- Institutional theory: -D over time
- Ecology: - D over time (selection), +D with resource changes
- Strategy: + D \rightarrow - competition, vice versa
 - Greckhamer, et al. (2008)
- Taxonomic and typologic studies of organizations
 - Basis for comparative science of organizations (McKelvey, 1982)
 - Fiss, 2007
-

QCA approach to diversity

- Create property space of theoretically most diverse possible set memberships
- Map cases onto property space to form truth table
- Analyze property space mapping to quantify diversity
 1. Are cases randomly assorted across property space of configurations?
 2. Observed diversity in comparison to possible diversity
 3. Concentration of cases in spaces within set of observed diversity
 4. Mapping function of simplest sets covering all observed cases
 - Simplified mathematical representation of complexity
 - Detect underlying patterns to diversity

Quantitative measures

- Reject hypothesis of random distribution
 - Chi-square test
- Ratio Observed:Possible
 - Proportion of observed to possible configurations
- Concentration
 - Any measure of concentration of observed cases across configurations (e.g., could even do Herfindahl index)
- Limited diversity
 - Reduce truth table using existence as outcome variable

Example: Weber's "iron cage" hypothesis

- Increasing formalization will remove all diversity in organization structure
- Bureaucracy is single configuration of six (6) dimensions: Hierarchy, Division of Labor, Formal Procedures, Written Rules, Technical Competence Required for Holding a Job, Difference in rewards across jobs
- Theoretically maximum diversity of configurations = 2^6 (64) sets that define property space of the six dimensions
- If true, "iron cage" hypothesis should find:
 1. Patterned: existing cases are not randomly occurring across 64 configurations.
 2. Limited diversity: most existing cases have membership only within a few of theoretically possible configurations
 3. Unity: most existing cases fall within a single configuration of having all six dimensions of a formal bureaucracy

Diversity mapping

- N=349 organizations from National Organizations Survey (Kalleberg, Knoke, Marsden, and Spaeth 1992)
 - Membership defined by median split of index of corresponding questions in the survey into high and low
- Truth table of the distribution of observed cases over terrain of 64 possible set membership configurations

Quantify diversity

- Not randomly assorted across property space
 - Chi-square = 393.88, 64. d.f. (expected = 349 cases/64 configurations = 5.453), $p=.001$
- Proportion observed configurations to possible is diverse
 - High: 45 of 64 possible locations on the map are filled (19 empty)
- Low concentration of cases in any configuration in property space
 - Low: Only 14 out of 45 have 10 or greater cases; none has 10% of cases

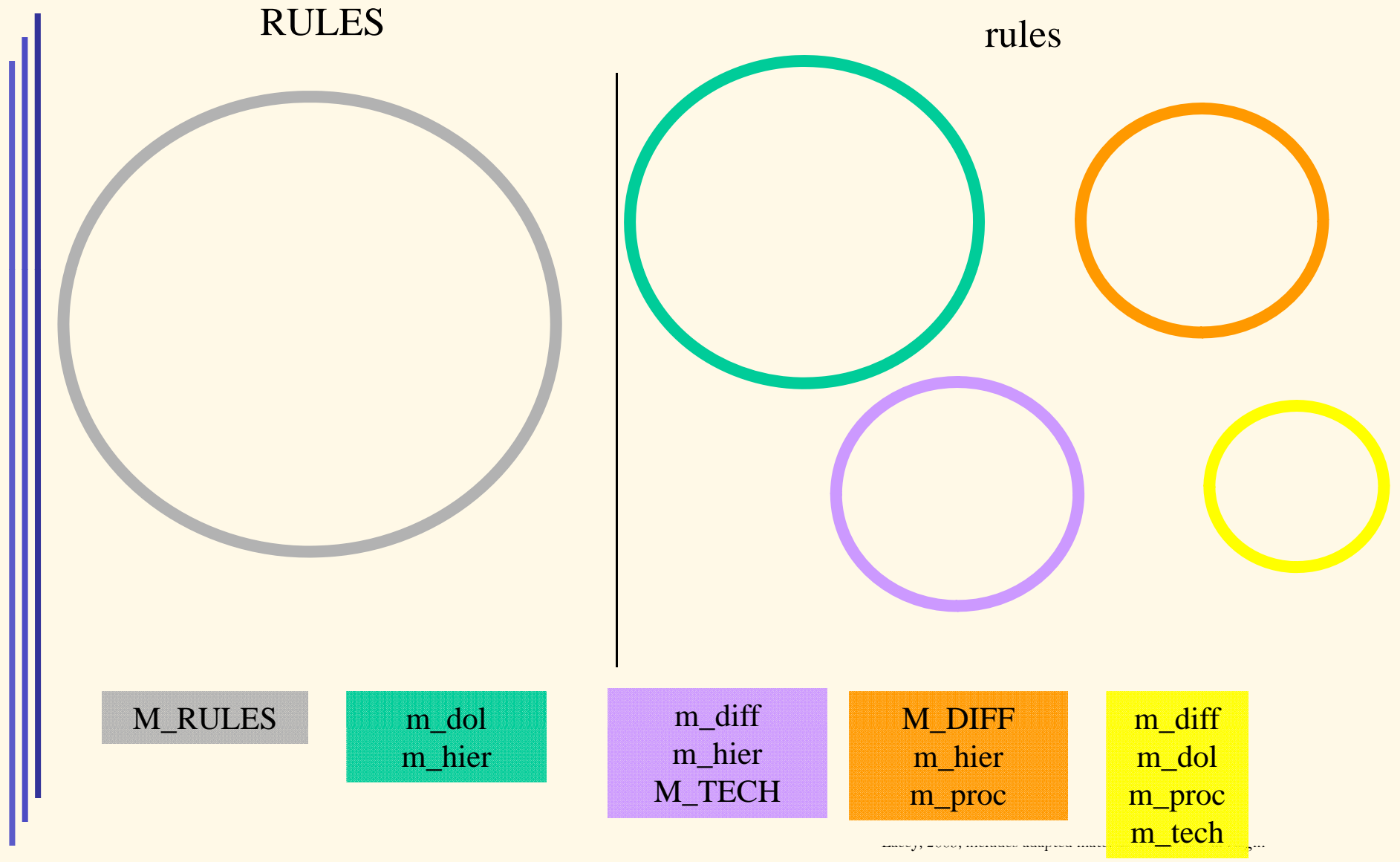
Capturing limited diversity

- Reduction of truth table (Table 2b)
- M_RULES + m_dol m_hier + m_diff m_hier M_TECH + M_DIFF m_hier m_proc + m_diff m_dol m_proc m_tech
- No configurations with contradictions included; so reduced for necessity and sufficient configurations only

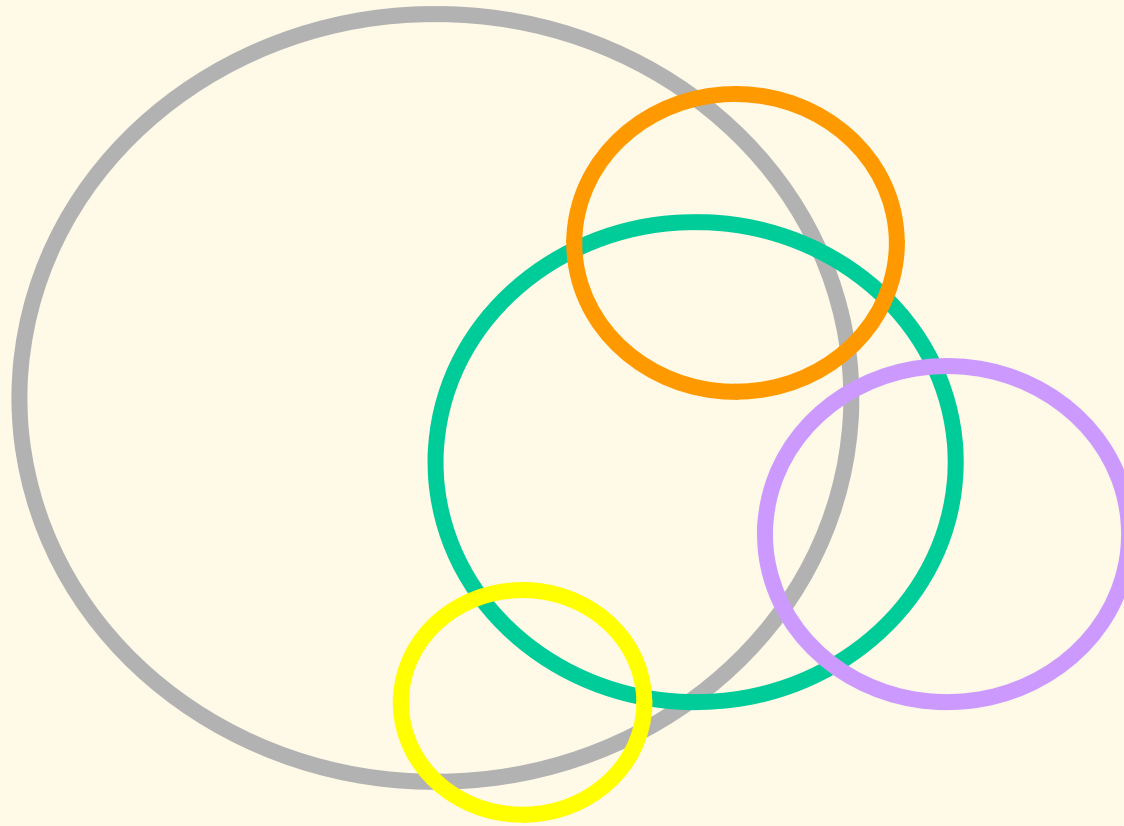
Diversity as set relationship

- (RULES) covers 71% of configurations (32 of 45)
- Set membership of configurations is mutually exclusive
 - All configurations with (RULES) separate from all configurations with (rules)
- Four terms cover the (rules) cases
- Set membership is not mutually exclusive; some cases belong to more than one term's set (sum > 349)
- All set membership overlaps with at least another set

No. of Map Locations (Configurations) Covered by Each Term in Boolean Equation (Table 2b)



Case (Organization) Membership Overlaps of Each Term in Boolean Equation Mapping Function (Table 2b)



M_RULES

m_dol
m_hier

m_diff m_hier
M_TECH

M_DIFF
m_hier
m_proc

m_diff
m_dol
m_proc
m_tech

Conclusions

- STM and QCA have multiple theoretical and practical layers at work simultaneously
- Emphasis on different layers for different purposes, but all three work at same time
- Larger opportunity of STM is not as complement to linear methods for small-N analyses, but as a new ontological and epistemological approach