

QUALITATIVE COMPARATIVE ANALYSIS

**Charles C. Ragin
Department of Sociology
University of Arizona
Tucson, AZ 85718**

<http://www.fsqca.com>

<http://www.compass.org>

<http://www.u.arizona.edu/~cragin>

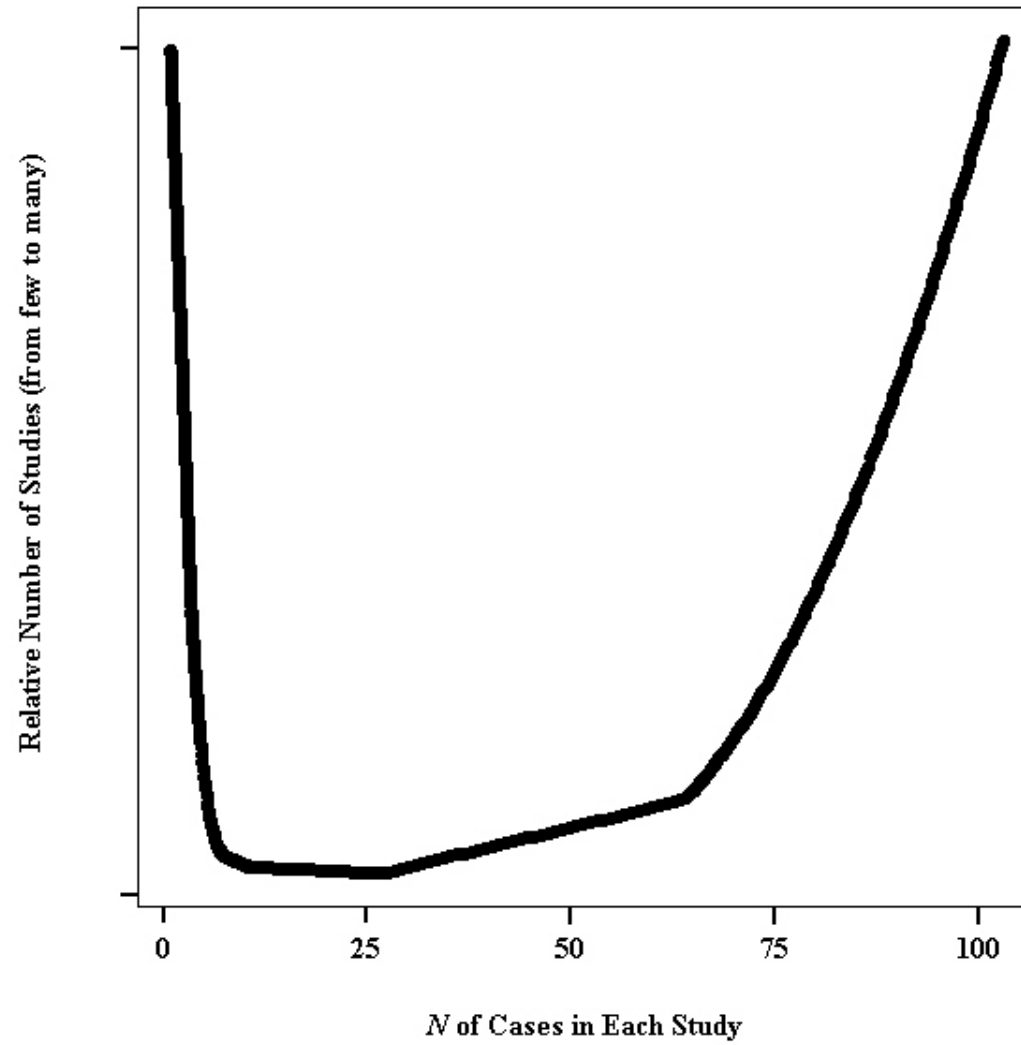
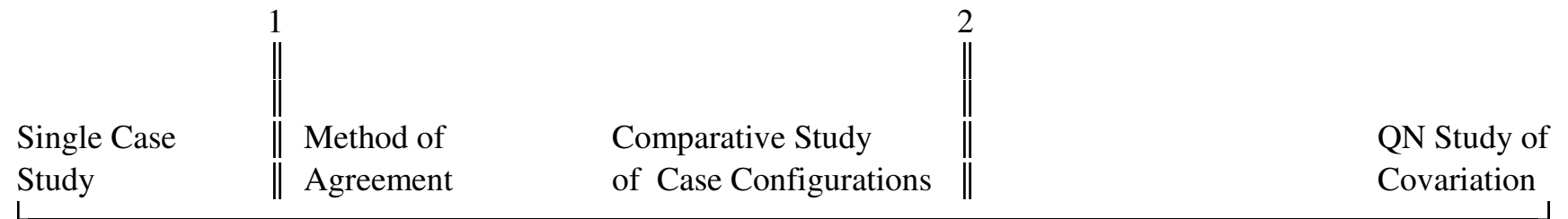


Figure 1.1 Plot of Relative Number of Studies against N of Cases in Each Study

Why the U-Shaped Pattern?

1. The nature of the discipline, including training, publishing, invisible colleges, and so on. Researchers tend to use methods they learn in graduate school, where training typically is bifurcated.
2. The once common belief that qualitative work is only for the mathematically impaired (elitism of quantitative researchers).
3. The underdevelopment of methods for medium-sized N s.
4. The difficulty of knowing a large number of cases in an an-depth manner.
5. The difficulty of keeping track of $(N)(N-1)/2$ paired comparisons.
6. The difficulty of considering 2^k logically possible combinations of conditions (relevant to counterfactual analysis), where k is the number of causal conditions.

THE CASE-ORIENTED/VARIABLE-ORIENTED DIMENSION



Case-Oriented
Small-N
Qualitative
Intensive
With-in Case Analysis
Problem of Representation

Variable-Oriented
Large-N
Quantitative
Extensive
Cross-Case Analysis
Problem of Inference

Some Assumptions:

1. Social scientists seek generalizations. They are interested in constructing statements about general patterns.
2. Cross-case analysis is central to the process of constructing generalizations. It is not a necessary ingredient, but is a very common way of arriving at general statements. (This assumption begs the question: What is a case?)
3. The results of cross-case analysis can be very misleading. The spurious correlation is the best known example of the limitations of cross-case analysis.
4. The best way to address the limitations of cross-case analysis is by complementing it with within-case analysis. If possible, it is good to balance cross-case and within-case analysis in social research. (This is a more precise version of the common admonition to combine qualitative and quantitative analysis.)
5. Causal processes are best observed at the level of the single case, through in-depth research.

The Middle Path

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
1. Proximate goals	Case study researchers focus on the problem of making sense of a very small number of cases, usually one and rarely more than three, selected because they are substantively or theoretically important in some way. The key concern is the representation of the case.	Comparative researchers study substantively or theoretically defined categories of cases (usually five to 50 or more), with the goal making sense of both individual cases and clusters of similar cases in the light of knowledge of cross-case patterns, and vice versa.	Variable-oriented research seeks to document general cross-case relationships between variables characterizing a large population of generic observations. The key focus is on the relative conformity of cross-case relationships with theoretically based models.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
2. Populations	<p>The case-study researcher's answer to "What is my case a case of?" may change throughout the course of the investigation, as the investigator learns more about the phenomenon in question and refines his or her guiding concepts and analytic schemes. The fact that a single case can be defined in multiple ways is usually seen as a strength, making the case "rich."</p>	<p>In comparative research, the investigator constructs a carefully delimited set of cases, using theoretical and substantive knowledge as guides. The boundary around this set is initially flexible and becomes more fixed as the research proceeds, through the interaction of ideas and evidence. Concept formation and empirical categorization go hand-in-hand.</p>	<p>In variable-oriented research, cases and populations are typically seen as given. The ideal-typic case (or "observation") is the survey respondent. Macrolevel cases such as countries are treated in the same generic manner. The key issue is how to derive a representative sample from the abundant supply of "given" observations.</p>

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
3. N of cases	<p>Case-study research is often defined by its focus on phenomena that are of interest because they are rare--that is, often an <u>N</u> of only one. Empirical depth is more important than breadth; therefore, enlarging the <u>N</u> is typically viewed as hazardous. Comparability of cases is never assumed and usually viewed as limited at best.</p>	<p>Comparative researchers often make strategic comparisons and thus need diverse cases. At the same time, they need to maintain case homogeneity because their cases should all be instances of or candidates for the same outcomes. Thus, comparative researchers must balance conflicting pressures when delimiting the set of relevant cases.</p>	<p>Variable-oriented researchers are encouraged to enlarge their number of cases whenever possible; more is always better. With more cases, researchers can make more precise estimates of the strength of the connections among variables. The individuality of each case is relegated to the error vector, giving the researcher a distilled representation of what is general across cases.</p>

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
4. Role of Theory	<p>Case-study researchers use in-depth study of cases to advance theory. Thus, they often choose cases that are anomalous in some way from the viewpoint of current theory. A case study is successful even if it succeeds in showing only that existing theory is inadequate. Thus, case selection is critically important.</p>	<p>Existing theory is rarely well-formulated enough to provide explicit hypotheses in comparative research. The primary theoretical objective of comparative research is not theory testing, but concept formation, elaboration, and refinement, and also theory development. Sharpening the definition of the set of relevant cases is often an important theoretical advance in itself.</p>	<p>In variable-oriented research, it is often presumed that researchers have well-defined theories and well-formulated hypotheses at their disposal from the very outset of their research. Theory testing is the centerpiece of social research. The ideal variable-oriented investigation adjudicates between competing theories.</p>

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
5. Conception of outcomes	<p>Case-study researchers often select cases specifically because of their uncommon or anomalous outcomes. The usual goal is to resolve the anomaly in a theoretically progressive way, based on in-depth knowledge of the selected case(s). Sometimes there is no sharp separation of causal conditions and outcomes, for an outcome may seem inherent in the constitution of the case.</p>	<p>Comparative researchers often begin by intentionally selecting cases that do not differ greatly from each other with respect to the outcome that is being investigated; they are all "positive cases." The constitution and analysis of the positive cases is usually a prerequisite for the specification of relevant negative cases--if they can be reasonably identified.</p>	<p>Variable-oriented researchers are advised to direct their attention to "dependent variables" that display a healthy range of variation across a systematic sample of cases drawn from a large population. Usually, the more fine-grained this variation, the better. Outcomes that do not vary across cases cannot be studied because there is no variation to explain.</p>

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
6. Causation	Case-study researchers examine causation holistically, in terms of a convergence of structures, actors, and events. They are also centrally concerned with sequences and timing of events, with an eye toward turning points and path dependence.	Comparative researchers usually look at causation in terms of multiple pathways. Positive cases often can be classified according to the general path each traveled to reach the outcome. Each path, in turn, can be seen as involving a different combination of relevant causal conditions.	Variable-oriented researchers assess the relative importance of competing independent variables in order to test theory. The key focus is on the relative importance of causal variables across cases, not on how they come together or combine in any single case. A single causal model is derived that applies equally to all cases.

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
<p>7. Within versus cross-case analysis</p>	<p>Case-study research is focused almost entirely on within-case patterns. Researchers examine parts of the case as mutually constitutive of each other and the whole they form together. Case-study researchers often ask: 'What kind of whole has parts like this?' as they explore connections among case aspects.</p>	<p>Comparative researchers focus on configurations of causally relevant characteristics of cases, with the goal of determining how relevant aspects fit together. They use cross-case analysis to strengthen and deepen within-case analysis, and vice versa. To the extent possible, comparative researchers try to balance cross-case and within-case analysis.</p>	<p>Variable-oriented researchers give priority to cross-case patterns. Aspects of cases are viewed primarily in terms of how they vary and co-vary across cases. How aspects of cases connect within each case is more or less ignored. The idiosyncrasies of cases cancel each other out, as deviations from general patterns are assigned to the error vector of probabilistic models.</p>

SETS ARE CENTRAL TO SOCIAL SCIENTIFIC DISCOURSE

Many, if not most, social scientific statements, especially empirical generalizations about cross-case patterns, involve set-theoretic relationships:

- A. Religious fundamentalists are politically conservative. (Religious fundamentalists are a subset of politically conservative individuals.)
- B. Professionals have advanced degrees. (Professionals are a subset of those with advanced degrees.)
- C. Democracy requires a state with at least medium capacity. (Democratic states are a subset of states with at least medium capacity.)
- D. "Elite brokerage" is central to successful democratization. (Instances of successful democratization are a subset of instances of elite brokerage.)
- E. "Coercive" nation-building was not an option for "late-forming" states. (States practicing coercive nation-building are a subset of states that formed "early.")

Usually, but not always (e.g., D), the subset is mentioned first. Sometimes, it takes a little deciphering to figure out the set-theoretic relationship, as in E.

CONVENTIONAL VIEW OF SETS

- Sets are binary, nominal-scale variables, the lowest and most primitive form of social measurement.
- The cross-tabulation of two sets is the simplest and most primitive form of variable-oriented analysis.
- This form of analysis is of limited value because: (1) the strength of the association between two binary variables is powerfully influenced by how they are created (e.g., the choice of cut-off values), and (2) with binary variables researchers can calculate only relatively simple measures of association. These coefficients may be useful descriptively, but they tell us little about the contours of relationships.
- In short, examining relations between binary variables might be considered adequate as a descriptive starting point, but this approach is too crude to be considered *real* social science.

Correlational Connections

- Correlation is central to conventional quantitative social science. The core principle is the idea of assessing the degree to which two series of values parallel each other across cases.
- The simplest form is the 2x2 table cross-tabulating the presence/absence of a cause against presence/absence of an outcome:

	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

- Correlation is strong (and in the expected direction) when there are as many cases as possible in cells #2 and #3 (both count in favor of the causal argument, equally) and as few cases as possible in cells #1 and #4 (both count against the causal argument, equally).
- Correlation is completely symmetrical.

Correlational Versus Explicit Connections

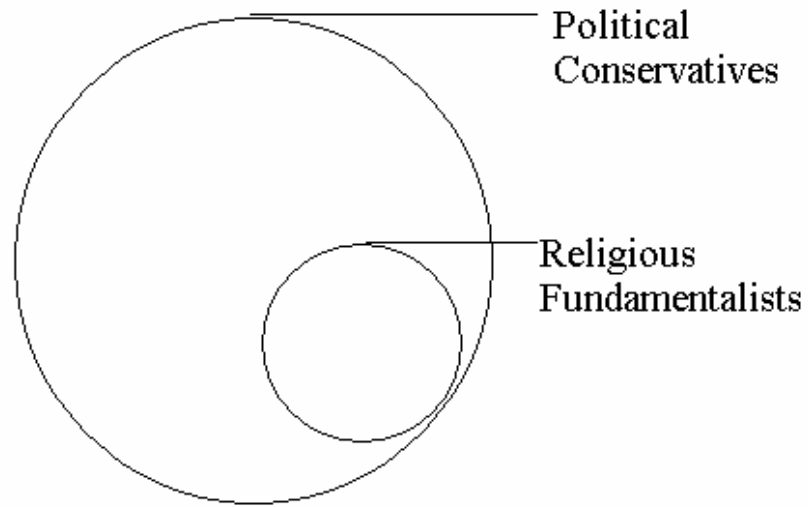
- A correlational connection is a description of tendencies in the evidence:

	Presidential form	Parliamentary form
3 rd wave democracy survived	8	11
3 rd wave democracy collapsed	16	5

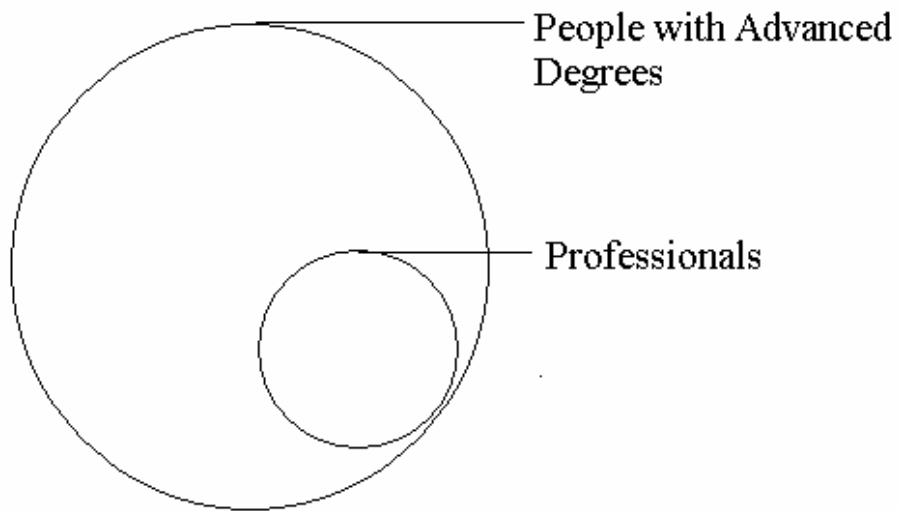
- An explicit connection is a subset relation or near-subset relation:

	Presidential form	Parliamentary form
3 rd wave democracy survived	18	16
3 rd wave democracy collapsed	6	0

In the second table all democracies with parliamentary systems survived, that is, they are a subset of those that survived. The first table is stronger and more interesting from a correlational viewpoint; the second is stronger and more interesting from the perspective of explicit connections.



The cause is a subset of the outcome.



The outcome is a subset of the cause.

NECESSITY AND SUFFICIENCY AS SUBSET RELATIONS

Anyone interested in demonstrating necessity and/or sufficiency must address set-theoretic relations. Necessity and sufficiency cannot be assessed using conventional quantitative methods.

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

SUFFICIENCY (WITHOUT NECESSITY)

I. Expressed as a simple truth table:

Cause	Outcome
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.

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).

Adding Causal Variables in Variable-Oriented Research

If the effects of two variables are additive, then the highest average level or probability of the outcome should occur when both causes are present, while the lowest should occur when both causes are absent.

	Neither parliamentary form nor multiparty	Only one of the two attributes present	Both parliamentary form and multiparty
3rd wave democracy survived	5	8	8
3rd wave democracy collapsed	10	7	3

ELABORATING EXPLICIT CONNECTIONS: OUTCOME IS A SUBSET OF THE CAUSE

When the goal is to establish that the outcome is a subset of a causal condition, the objective is to move cases from cell 1 to cell 2 (i.e., to drain cell 1 of cases and thereby establish an explicit connection). In effect, the causal argument must be made more inclusive, which can be accomplished using logical *or*. Generally, this use of logical *or* entails moving up the ladder of abstraction to a more general conceptualization of the causal condition or construct.

For example, to survive as a third-wave democracy it might be necessary to have EITHER a parliamentary form OR a multiparty system. At a more abstract level, these two conditions could be seen as substitutable ways of accomplishing political inclusiveness, which in turn could be interpreted as a necessary condition for democratic survival.

ILLUSTRATION OF USE OF LOGICAL *OR* TO IDENTIFY EXPLICIT CONNECTION

	X Absent	X Present			both absent	X <i>or</i> Z Present
Outcome Present	5	25	←before after→	Outcome Present	0	30
Outcome Absent	15	15		Outcome Absent	12	18

By identifying a substitutable causal condition (and moving to a more general conceptualization), it is possible to identify an explicit connection--the outcome is now a subset of the reconstituted cause.

It is important to understand that there is a dialogue of ideas and evidence in this procedure. The goal is not simply to find a causal condition that improves fit with the outcome, but to use ideas to craft a more encompassing causal condition.

Logical *or* is central to the process of trying to empty cell 1 of cases. In some instances, the encompassing causal condition may be interpreted as a necessary condition.

ELABORATING EXPLICIT CONNECTIONS: THE CAUSE IS A SUBSET OF THE OUTCOME

- When the goal is to establish that the cause is a subset of the outcome, the goal is to move cases from cell 4 to cell 3 (i.e., to empty cell 4 of cases and thereby establish an explicit connection). In effect, the causal argument must be made more restrictive, which is accomplished through the use of logical *and*. Generally, this use of logical *and* also entails moving toward a more nuanced conceptualization of the causal conditions.
- For example, one recipe for survival as a 3rd wave democracy might be to combine a party system that permits representation of minorities (i.e., a multiparty system) with an institutional form that fosters coalition building and political bargaining (i.e., the parliamentary form). The set of cases combining these two traits might constitute a subset of those that survive.

The effort to empty cell 4 of cases is connected with greater theoretical nuance and specificity regarding the nature of the causal mechanisms.

	X Absent	X Present			X or Z absent	X*Z Present
Outcome Present	16	14	←before after→	Outcome Present	18	12
Outcome Absent	24	6		Outcome Absent	30	0

Note that when seeking to empty cell 4 of cases, the goal is to refine one of perhaps several recipes for the outcome.

CAUSAL COMPLEXITY

Another important benefit of set theoretic analysis is that it is much more compatible with the analysis of causal complexity than conventional techniques.

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

Possible findings include:

(1) technology \rightarrow strikes

(2) technology*wages \rightarrow strikes

(3) technology + wages \rightarrow strikes

(4) technology*wages + overtime*sourcing \rightarrow strikes

In (1) technology is necessary and sufficient; in (2) technology is necessary but not sufficient; in (3) technology is sufficient but not necessary; in (4) technology is neither necessary nor sufficient. The fourth is the characteristic form of causal complexity: no cause is either necessary or sufficient.

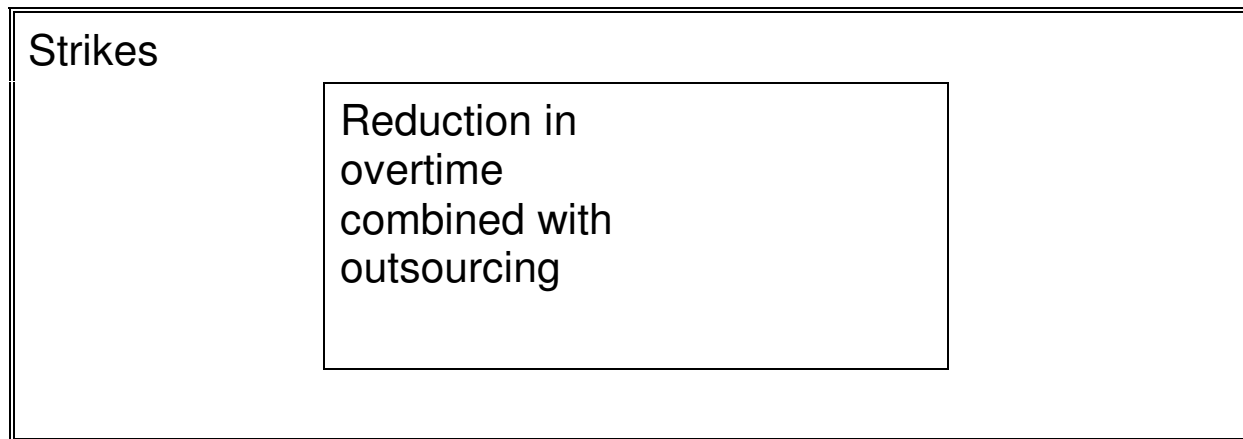
ASSESSING CAUSAL COMPLEXITY

I. Logical equation: technology*wages + overtime*sourcing → strikes

II. Formulated as a partial crosstabulation:

	Causal combination absent	Causal combination present
Strike present (1)	Cell 1: 20 cases	Cell 2: 23 cases
Strike absent (0)	Cell 3: 18 cases	Cell 4: 0 cases

III. Expressed as a Venn diagram:



The key to assessing the sufficiency of a combination of conditions, even if it is one among many combinations, is to select on instances of the combination and assess whether these instances agree on the outcome.