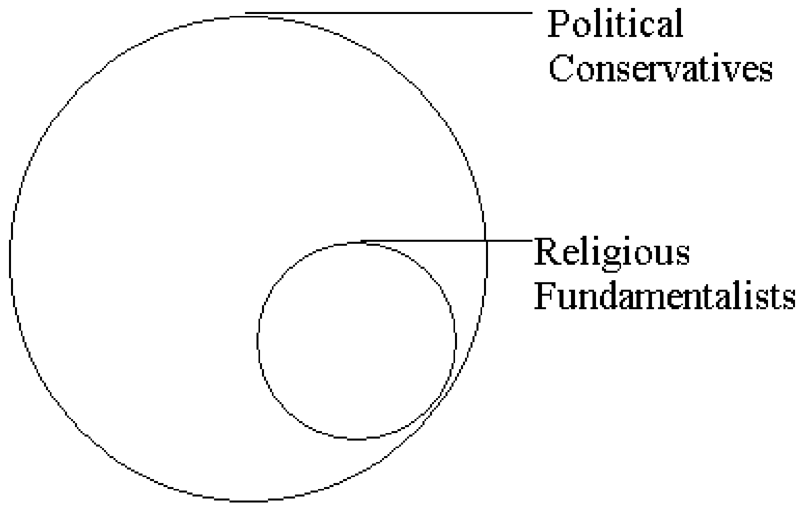
SETS ARE CENTRAL TO SOCIAL SCIENTIFIC DISCOURSE

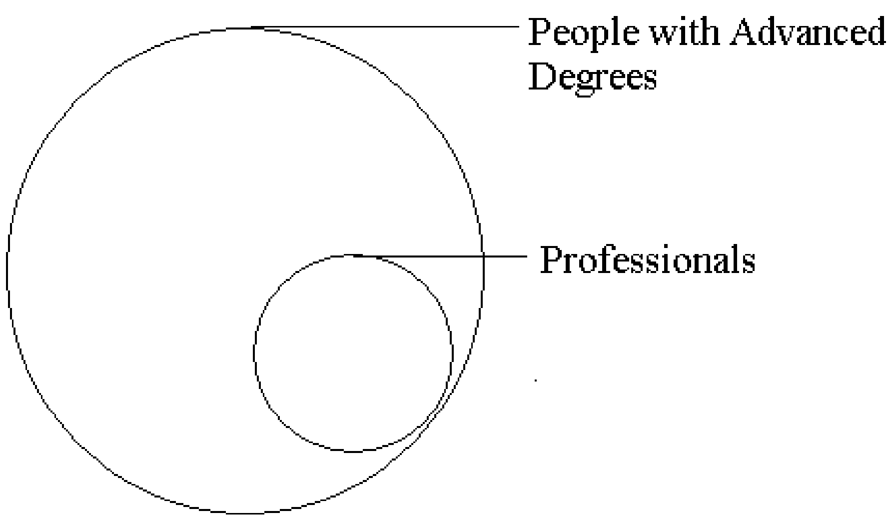
Many, if not most, social scientific statements, especially empirical generalizations about cross-case patterns, involve set-theoretic relationships:

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.")
F. “No bourgeoisie, no democracy.” (Barrington Moore, Jr.) An urban bourgeoisie is one of the preconditions for the development of democracy.

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.



The cause is a subset of the outcome.



The outcome is a subset of the cause.

Explaining Outcomes and Set-Theoretic Reasoning

1. When explaining outcomes, the focus is on temporally bounded qualitative change.

2. Often, the explanation of the outcome is combinatorial, citing a confluence of actors, events, and structures.

3. Explaining “how things happen(ed)” is central to case study and comparative research: “The origins of _________.” “Becoming a _________.” “The decline of ________.” “The disappearance of ___________.” “The emergence of ___________.”

4. With more than one case, there may be multiple combinations of conditions (i.e., multiple causal recipes), with different cases following different paths to the same outcome.

5. Each “path” can be understood as a different combination of conditions, which in turn can be formulated as an intersection of sets (a causal recipe).

6. Sufficiency is established when it can be shown that instances of the causal recipe constitute a subset of instances of the outcome.

CONVENTIONAL VIEW OF SETS

- Sets are binary, nominal-scale variables, the lowest and most primitive form of social measurement.

- The cross-tabulation of two binary 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 Set-Theoretic Connections

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

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

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

	Presidential form	Parliamentary form
3rd wave democracy survived	18	16
3rd 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.

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) ≤ (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).

NECESSITY (WITHOUT SUFFICIENCY)

I. Expressed as a simple truth table:

Cause	Outcome
1	1
1	0
0	0

II. Expressed as an inequality:

(value of the outcome) ≤ (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). This strategy is central to many forms of qualitative research.

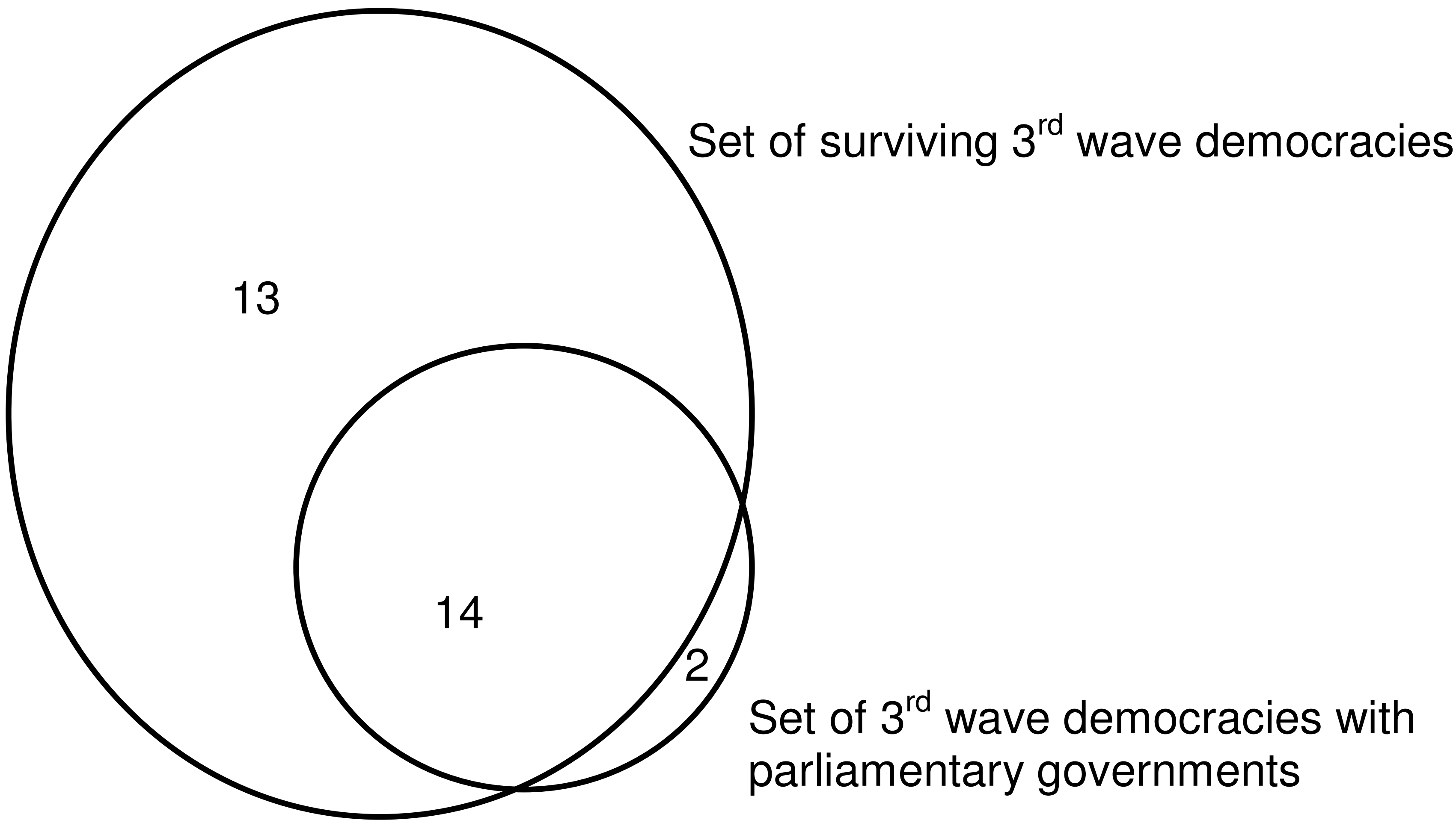
The Degree of Consistency of a Subset Relation

The consistency of a crisp subset relation is straightforward. It is simply the number of cases displaying both the cause and the outcome (cell 2) divided by the number of cases displaying the causal condition (the sum of cells 2 and 4). For example, in the following table the consistency of the evidence with the argument that having a parliamentary form of government is sufficient for the survival of a 3rd wave democracy is:

$$14/(2 + 14) = 14/16 = 0.875$$

	Presidential form	Parliamentary form
3rd wave democracy survived	13	14
3rd wave democracy collapsed	11	2

Only cases in the second column are involved in the assessment of consistency of the evidence with the subset relation. The following is a Venn diagram illustrating this high but less-than-perfect degree of consistency.



Consistency of subset relation = 14/(2+14) = 14/16 = 0.875

The Degree of Coverage of a Subset Relation

Once it has been established that a subset is at least roughly consistent, it is possible to assess its degree of coverage. The calculation of coverage for crisp sets is straightforward. It is simply the number of cases with both the causal condition and the outcome (again, cell 2) divided by the number of cases with the outcome (cell 1 + cell 2). In other words, coverage answers the question: What proportion of cases with the outcome has been “explained”? or How common is the cause (or causal combination) among the cases with the outcome?

$$14/(13 + 14) = 14/27 = 0.519$$

	Presidential form	Parliamentary form
3rd wave democracy survived	13	14
3rd wave democracy collapsed	11	2

Only cases in the first row are involved in the assessment of coverage. The previous Venn diagram also illustrates the concept of coverage. The key consideration is the proportion of the larger set covered by the interior set.

Set Coincidence

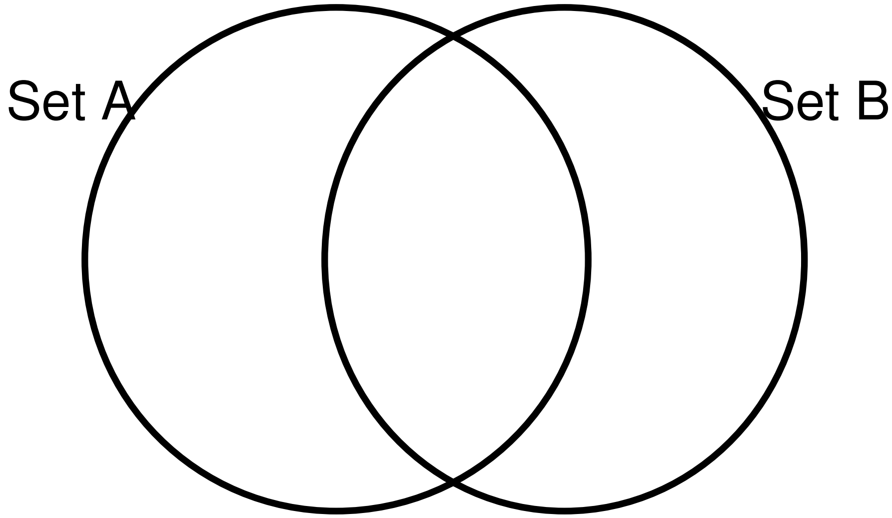
Set coincidence combines and bridges consistency and coverage. Set coincidence focuses on the degree to which two sets overlap—that is, the degree to which they are one and the same set.

While degree of set coincidence can be assessed using multiple sets (i.e., more than two), it is easiest to grasp the basic principles using two sets. For example, the degree to which the set of *surviving* 3^{rd} wave democracies and the set of 3^{rd} wave democracies *with parliamentary governments* are "one and the same" is indicated by the degree to which the cases than have *both* of these two traits embraces the set of cases that have *either* trait. In other words, set coincidence is the number of cases found in the intersection of two sets, expressed relative to the number of cases found in their union:

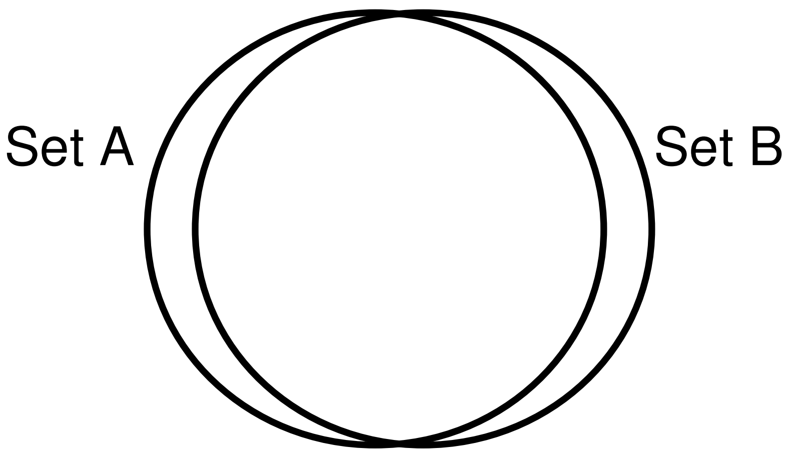
(# of cases in intersection)/(# of cases in union)

In the table shown on the previous slide, the coincidence of "Parliamentary" and "Democracy Survived" is 14/29 = 0.483 (i.e., relatively modest).

Here are two graphic examples, showing the contrast between high and low coincidence:



Low set coincidence: set intersection is a small fraction of set union.



High set coincidence: set intersection almost fully “covers” set union.