CHAPTER 7. COMPARATIVE ANALYSIS WITH BOOLEAN ALGEBRA

This chapter is divided into three main parts. First, the five methods for testing single factors as sufficient and/or necessary conditions are presented. These are based on the Methods of Agreement and Difference, and on the Joint Method (or Indirect Method). Second, combinatorial methods are introduced in which explanation is based on the configuration of the values of independent variables. Both these sections deal with dichotomous data, that is, variables that assume either a value of presence (1) or absence (0). Third, the methods are extended to nondichotomous data, namely to fuzzy-set analysis.

The Search for Sufficient Conditions

Based on Effects (Method 1)

As seen earlier, sufficient conditions are easier to interpret in terms of sufficient "causes" than necessary conditions. We therefore start with sufficient conditions. If *C* (a hypothetical cause) is a sufficient condition for *E* (a hypothetical effect), then *C* implies *E* and the conditional truth table looks as follows:

	C	→	*E*
I	1	1	1
II	1	0	0
III	0	1	1
IV	0	1	0

with $C \rightarrow E$ (the 0/1 values in the cells never change).

If *C* is a sufficient condition for *E*, then there is never a case in which *C* is present and *E* is absent. That is to say, there is never the combination $C = 1$ and $E = 0$ (combination II) or in Bayesian probability notation $P(C \mid {\sim}E) = 0$. Note, therefore, that it is combination II that rejects the hypothesis. If, for example, PR electoral systems are a sufficient condition for multiparty systems, then there must never be a two-party system when PR exists.

According to the deductive argument:

If C is a sufficient condition for E, then C cannot occur in the absence of E
There is one (or more than one) instance in which C occurs in the absence of E

$\therefore$ C is not a sufficient condition for E

In practice, with this research strategy, *cases are selected such that all $E = 0$*. The potential sufficient conditions (C_1, C_2, C_3, . . .) are then examined and, possibly, excluded. According to combination II of the truth table, we eliminate the combinations $C = 1$, $E = 0$.

Instances (Cases)	*Potential Sufficient Conditions (Independent Variables)*										*Effect (Dependent Variable)*
	C_1	C_2	C_3	C_4	C_5	.	.	.	.	C_m	E
1	1	0	1	1	0	.	.	.	.	.	0
2	0	0	1	1	0	.	.	.	.	.	0
3	0	0	1	1	0	.	.	.	.	.	0
4	1	0	0	1	0	.	.	.	.	.	0
5	1	0	0	1	1	.	.	.	.	.	0

According to this table, we exclude as sufficient causes all potential conditions with the exception of C_2. That is to say, we do not reject cases in which $C = 0$, $E = 0$ (combination IV).

Based on Causes (Method 2)

If C is a sufficient condition for E, every time that C is present E must also occur. If E is not present, then C is not a sufficient condition for E. Again, C implies E ($C \rightarrow E$) and the conditional truth table is the same as for Method 1 above.

As for methods based on effects, if C is a sufficient condition for E, then there is never a case in which C is present and E is absent. That is to say, that there is never the combination $C = 1$ and $E = 0$ (combination II) or in Bayesian probability notation $P(C \mid \sim E) = 0$. Note, again, that it is combination II that rejects the hypothesis.

The deductive argument too remains the same:

If *C* is a sufficient condition for *E*, then *C* cannot occur in the absence of *E*
There is one (or more than one) instance in which *C* occurs in the absence of *E*

∴ *C* is not a sufficient condition for *E*

In practice, however, the research strategy is different. Rather than selecting cases in which $E = 0$, that is, negative cases in which the effect does not occur, we select cases according to the value $C = 1$, that is, all cases in which the condition we hypothesize as sufficient is present. According to this method, *cases are selected such that all C = 1*. The effects (E_1, E_2, E_3, . . .) are then examined and, possibly, alternative hypotheses are excluded. According to combination II of the truth table, we eliminate the combinations $C = 1$, $E = 0$.

Instances (Cases)	*Potential Sufficient Condition (Independent Variable)*	*Effects (Dependent Variables)*							
	C_1	E_1	E_2	E_3	E_4	.	.	.	E_m
1	1	0	1	1	0	.	.	.	.
2	1	0	1	1	0	.	.	.	.
3	1	0	1	1	0	.	.	.	.
4	1	0	0	1	1	.	.	.	.
5	1	0	0	1	0	.	.	.	.

This is a more "experimental" and practice-oriented approach in which one controls the causes and tries to identify what their effects are. According to this table, we exclude *C* as a sufficient condition for E_1, E_2, and E_4. We do not exclude *C* as a sufficient condition of E_3. That is to say, we do not reject cases in which $C = 1$, $E = 1$ (combination I).

The Search for Necessary Conditions

Based on Effects (Method 3)

With this method, given an event *E*, we want to know which factors, among a number of potential alternative necessary conditions, are to be rejected and which not. As seen above, if *C* (a hypothetical cause) is a necessary condition for *E* (a hypothetical effect), then *C* is implied by *E* ($C \leftarrow E$) and the conditional truth table looks as follows:

	E	→	*C*
I	1	1	1
II	1	0	0
III	0	1	1
IV	0	1	0

with $E \rightarrow C$ rather than $C \rightarrow E$ (as for sufficient conditions).

If *C* is a necessary condition for *E*, then there is never a case in which *C* is absent and *E* is present. That is to say that there is never the combination $C = 0$ and $E = 1$ (combination II) or in Bayesian probability notation $P(\sim C \mid E) = 0$. It is again combination II that rejects the hypothesis. If a civic political culture is a necessary condition for the stability of democracies, then there is never a case in which a civic political culture is absent and democracy is stable.

According to the deductive argument:

If *C* is a necessary condition for *E*, then *C* cannot be absent in the presence of *E*
There is one (or more than one) instance in which *C* is absent in the presence of *E*

∴ *C* is not a necessary condition for *E*

In practice, with this research strategy, *cases are selected such that all E* = 1. The potential necessary conditions (C_1, C_2, C_3, . . .) are then examined and, possibly, excluded. According to combination II of the truth table, we eliminate combinations in which $C = 0$, $E = 1$.

Instances (Cases)	*Potential Necessary Conditions (Independent Variable)*										*Effect (Dependent Variables)*
	C_1	C_2	C_3	C_4	C_5	.	.	.	.	C_m	*E*
1	1	0	1	1	0	.	.	.	.	.	1
2	0	0	1	1	0	.	.	.	.	.	1
3	0	0	1	1	0	.	.	.	.	.	1
4	1	0	0	1	0	.	.	.	.	.	1
5	1	0	0	1	1	.	.	.	.	.	1

According to this table, we exclude as necessary conditions for E all potential conditions with the exception of C_4. That is to say, we do not reject cases in which $C = 1$, $E = 1$ (combination I).

As Braumoeller and Goertz (2000, p. 846) note, when testing the proposition that C is a necessary condition for E if C is always present when E occurs, cases in which $E = 0$ are irrelevant. This appears in the following table:

		C 0	*C* 1
E	0	—	—
	1	0	1

In this table, C is always present when E is present: $P(C \mid E) = 1$. We thus select only cases in which $E = 1$.

Based on Causes (Method 4)

If C is a necessary condition for E, every time that C is absent E cannot occur. If E is present, then C is not a necessary condition for E. Again, C implies E ($C \rightarrow E$) and the conditional truth table is the same as above in Method 3 based on effects.

As for methods based on effects, if C is a necessary condition for E, then there is never a case in which C is absent and E is present. That is to say, that there is never the combination $C = 0$ and $E = 1$ (combination II) or in Bayesian probability notation $P(\sim C \mid E) = 0$. It is always combination II that rejects the hypothesis.

The deductive argument too remains the same:

If C is a necessary condition for E, then C cannot be absent in the presence of E
There is one (or more than one) instance in which C is absent in the presence of E

$\therefore$ C is not a necessary condition for E

In practice, however, the research strategy is different. Rather than selecting cases in which $E = 1$, that is, positive cases in which the effect occurs, we select cases according to the value $C = 0$, that is, all cases in which the condition we hypothesize as necessary is absent. With this research strategy *cases are selected such that all* $C = 0$. The effects (E_1, E_2, E_3, . . .) are then examined and, possibly, alternative hypotheses are excluded. According to combination II of the truth table, we eliminate combinations in which $C = 0$, $E = 1$.

Instances (Cases)	*Potential Necessary Condition (Dependent Variable)*	*Effects (Independent Variables)*							
	C_1	E_1	E_2	E_3	E_4	.	.	.	E_m
1	0	1	1	0	0	.	.	.	.
2	0	1	1	0	0	.	.	.	.
3	0	0	1	0	0	.	.	.	.
4	0	0	0	0	1	.	.	.	.
5	0	0	0	0	1	.	.	.	.

Again, being a method based on causes, this is a more "experimental" approach. According to this table, we exclude *C* as a necessary condition for E_1, E_2, and E_4. We do not exclude *C* as a necessary condition of E_3. We do not reject cases in which $C = 0$, $E = 0$ (combination IV). Testing the proposition that *C* is a necessary condition for *E* if *E* does not occur in the absence of *C*, cases in which $C = 1$ are irrelevant. This appears in the following table:

		C = 0	*C* = 1
E	0	1	—
	1	0	—

C is never present when *E* does not occur: $P(\sim C \mid \sim E) = 1$. We select only cases in which $C = 0$.

In sum, in all four methods we reject *H* according to combination II (0,1). With the methods based on causes we do not reject *H* based on combination IV (0,0) and with the methods based on effects we do not reject *H* based on combination I (1,1).

These four methods are instruments for rejecting false hypotheses to be used with different research strategies. If we want to test if a PR electoral system is a *sufficient condition* for multiparty systems (MPS) we can (1) based on *causes* select cases in which PR = 1 and see if in all MPS = 1, and (2) based on *effects* select cases in which MPS = 0 (two-party systems) and see if there are cases in which PR = 1. Conversely, if we want to test if a PR electoral system is a *necessary condition* for MPS we can (1) based on *causes* select cases of majoritarian systems (PR = 0) and see if there is MPS and (2) based on *effects* select cases in which MPS = 1 and see if there are cases in which PR = 0.

In conclusion, the control of hypotheses can be done in various ways: select causes and observe their effects or select effects and track their causes. The choice of a research strategy often depends on which and how many cases are available. In practice, the combination of diverse methods always strengthens results.

"Trivialness"

Braumoeller and Goertz are the first to formalize "trivialness." They illustrate the point by asking "what makes gravity trivially necessary for war?" (2000, p. 854) (see also Goertz & Starr, 2003). There are two main forms of trivialness and one case of nontriviality:

1. *Trivial Type 1* (left-hand table). *C* is a necessary condition for *E* if *C* is always present when *E* occurs. If we select cases with $E = 1$, there must always be $C = 1$ (*Method 3*). However, *C* may be present even when $E = 0$. Is gravity a necessary condition for war? Yes, because gravity is always present ($C = 1$) when there is a war ($E = 1$), however it is also present when there is no war ($E = 0$). In this case there is no variation in the *independent* variable (*C*). Gravity is a trivial necessary condition for wars when it is present in both cases of war and nonwar:

$$P(C \mid \sim E) = 1 \text{ and } P(C \mid E) = 1$$

2. *Trivial Type 2* (center table). *C* is a necessary condition for *E* if *E* does not occur in the absence of *C*. If we select cases with $C = 0$, there must always be $E = 0$ (*Method 4*). However, *E* may be absent even when $C = 1$. Is the presence of at least one authoritarian state a necessary condition for war? Yes, because war is always absent ($E = 0$) when there are no authoritarian states ($C = 0$). However, *E* is also absent when there are authoritarian states. In this case there is no variation in the *dependent* variable (*E*). Authoritarian states are a trivial necessary condition for wars when there are no wars:

$$P(\sim C \mid \sim E) = 1 \text{ and } P(C \mid \sim E) = 1$$

Trivial Type 1

		C = 0	C = 1
E	0	0	1
	1	0	1

Trivial Type 2

		C = 0	C = 1
E	0	1	1
	1	0	0

Nontrivial

		C = 0	C = 1
E	0	1	0
	1	0	1

3. *Nontrivial* (right-hand table). To avoid Types 1 and 2 of trivialness there must be variation in both the *independent and dependent* variables (*C* and *E*). To avoid Trivial Type 1 we need a variation in the independent variable (*C*). To avoid Trivial Type 2 we need a variation in the dependent variable (*E*). In practice, this means that we need to use both *Methods 3* and *4* to assess for nontrivial necessary conditions. If PR is a necessary condition for MPS, PR is always present when MPS occur: $P(C \mid E) = 1$. However, to be nontrivial (Type 1), it must not be present when MPS are absent: $P(C \mid {\sim}E) = 0$ (rather than =1). In addition, if PR is a necessary condition for MPS, MPS does not occur in the absence of PR: $P({\sim}C \mid {\sim}E) = 1$. However, to be nontrivial (Type 2), PR must not occur when there are no MPS: $P(C \mid {\sim}E) = 0$ (rather than =1). In sum:

$$P(C \mid E) = 1 \text{ and } P({\sim}C \mid {\sim}E) = 1$$

Sufficient and Necessary Conditions (Method 5)

The four methods discussed above provide the basis for more complex analyses. First, they allow identification of conditions that are both sufficient and necessary (this section). Second, they provide the tools for multivariate analysis and compound statements (next section).

To identify conditions that are both sufficient and necessary one uses two truth tables. The combination of the truth table for sufficient conditions with the truth table for necessary conditions allows one to identify conditions that are both sufficient *and* necessary. However, instead of rejecting hypotheses uniquely on the basis of combination II, the combinations leading to the rejection of a hypothesis are two: combinations II and III.

The truth table looks as follows

	C	↔	*E*
I	1	1	1
II	1	0	0
III	0	0	1
IV	0	1	0

with ↔ (or ≡) symbolizing the *equivalence* or "*double*" *implication*. The difference with "simple" implication is that both combinations II and III have a 0 in the central column rather than combination II only.

In practice, the two truth tables are used subsequently. First, one eliminates conditions that are not sufficient. Second, among those that "survived" the first phase, one eliminates conditions that are not necessary. The remaining condition(s) are both sufficient *and* necessary.

Instances (Cases)	*Potential Sufficient and Necessary Conditions (Independent Variables)*										*Effect (Dependent Variable)*
	C_1	C_2	C_3	C_4	C_5	.	.	.	.	C_m	E
1	1	1	1	1	1	.	.	.	.	.	1
2	1	0	1	0	1	.	.	.	.	.	1
3	1	0	0	1	1	.	.	.	.	.	1
4	1	1	0	0	1	.	.	.	.	.	1
5	1	1	0	1	1	.	.	.	.	.	1
6	1	0	1	0	0	.	.	.	.	.	0
7	1	1	1	0	0	.	.	.	.	.	0

If C is a sufficient condition for E, then when C is present E is always present too: $P(C \mid E) = 1$; and when C is present E can never be absent: $P(C \mid \sim E) = 0$. Otherwise we reject H on the basis of combination II. In the table above this eliminates C_1, C_2, and C_3 as sufficient conditions for E. Furthermore, if C is a necessary condition for E, then when E is present, C must be present too: $P(C \mid E) = 1$; and when E is present C can never be absent: $P(\sim C \mid E) = 0$. Otherwise we reject H on the basis of combination III. This eliminates C_4 as a necessary condition for E. C_5 is the only sufficient and necessary condition in this example. This method is based on the Joint Method of Agreement and Difference (or Indirect Method).

Multivariate Analysis With Logical Algebra

The previous methods have sometimes been criticized for being complex and cumbersome in particular when applied to multivariate analysis.

Improvements came from the further development of logical calculus (Cohen & Nagel, 1934; Nagel, Suppes, & Tarski, 1963; Roth, 2004; von Wright, 1951). First, this section presents the basic operators of logical algebra and illustrates their use in multivariate research designs. Second, Boolean analysis is introduced. Recent influential contributions have stressed the possibilities set algebra offers for the study of multivariate relationships (Ragin, 1987, pp. 85–163). The possibility of taking into account explanatory variables as configurations or combinations is one of the main strengths of modern comparative research designs.

Compound Statements

In multivariate analysis researchers look for sufficient and/or necessary conditions in the form of *compound attributes* rather than simple attributes, that is, "packages" of attributes. Ragin has made this "combinatorial logic" the distinctive feature of the comparative method (Ragin, 1987, p. 15). Rather than testing the empirical validity of hypotheses concerning potential sufficient or necessary conditions one by one or in an additive logic, conditions are tested when they combine in specific ways. To understand how this works, a few basic elements of logical algebra must be introduced.

Multivariate analysis is based on three fundamental Boolean *operators* or *connectives* for compound statements: AND, OR, and NOT.

1. The *conjunction* (AND) symbolized as $\cdot$ (or $\wedge$). The conjunction produces a compound statement in which both components (C_1 and C_2) are true (present). Whenever either component (or both) is false (0), the conjunction is false. The truth table shows the value of a compound statement for all combinations of values for its components:

C_1	C_2	$C_1 \cdot C_2$
1	1	1
1	0	0
0	1	0
0	0	0

2. The (inclusive) *disjunction* (OR) symbolized as + (or $\vee$).[20] The disjunction produces a compound statement in which either (or both) its components is true. It is false only when both of them are false.

The truth table shows the value of a compound statement for every possible combination of values for its components:

C_1	C_2	$C_1 \vee C_2$
1	1	1
1	0	1
0	1	1
0	0	0

3. The *negation* (NOT) symbolized as ~. The negation produces a statement that reverses the truth value of any statement (simple or compound). It is particularly important as it represents the absence of a causal condition ($C = 0$) or outcome ($E = 0$).[21]

Take the following example of a *conjunction* (AND). We may find that the single attribute PR is not a sufficient condition for MPS to occur nor that the single attribute social fragmentation (SF) is a sufficient condition for MPS, but that the compound attribute "PR and SF" (PR • SF) is a sufficient condition to produce MPS. According to Method 2 we would eliminate both PR and SF singularly as sufficient conditions for MPS (in Case 4 PR does not produce a MPS and in Case 5 SF does not produce a MPS). However, the presence of both PR and SF is a sufficient condition for MPS.

Instances (Cases)	*Potential Sufficient Conditions*			*Effect*
	PR	*SF*	*PR • SF*	*MPS*
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	0	0	0
5	0	1	0	0

In Boolean algebra the conjunction AND is called *multiplication*. The product is a *specific combination of causal conditions*. The statement PR • SF → MPS is written as

$$\text{MPS} = \text{PR} \cdot \text{SF} \text{ (or, simply, MPS} = \text{PR SF)}$$

When PR and SF are both present, then MPS is also present. With this type of notation $1 \cdot 0 = 0$ or, conversely, $0 \cdot 1 = 0$. If only one of the two components is present, then the outcome does not occur. The presence of PR is combined with the presence of SF to produce MPS, whereas the absence of any of the two simple statements does not lead to the outcome. A compound statement of this sort can also include "absence" of properties. For example, it is only the combination of majoritarian electoral system (M), "no SF" (~SF), and "no territorially concentrated minorities" (~TCM) to produce a two-party system (TPS),

$$\text{TPS} = \text{M} \cdot \sim\!\text{SF} \cdot \sim\!\text{TCM} \text{ (or TPS = M sf tcm)}$$

with *uppercase* letters indicating the presence of the attribute and *lowercase* letters indicating the absence (negation) of the attribute.

In Boolean algebra the *disjunction* OR is called *addition* and is symbolized through +. Here the addition consists of the fact that if any of the conditional components is present, then the outcome occurs. In this type of algebra therefore, $1 + 1 = 1$. If, for example, we ask what leads to a loss of votes for a party at a given election (LV), we may find evidence that various factors lead to the same outcome: a poor performance in government (PP), the emergence of a new concurrent party in the same ideological family (NP) or a political scandal involving the leader of the party (PS). If any one, or any two or all three factors are true, then the outcome LV occurs. The conditional statement $\text{PP} \vee \text{NP} \vee \text{PS} \rightarrow \text{LV}$ is written as

$$\text{LV} = \text{PP} + \text{NP} + \text{PS}$$

meaning that either a poor performance in government or a new concurrent party or a political scandal can each cause a loss of votes for the party (or all three or any two).

Compound statements based on conjunctions and disjunctions are crucial to distinguish: (1) sufficient but not necessary conditions, (2) necessary but not sufficient conditions, (3) neither sufficient nor necessary conditions, and (4) both sufficient and necessary conditions.

Disjunctions and Multiple Causation

1. *Sufficient but not necessary conditions*. Disjunctions or Boolean additions (OR) are particularly important because they allow to formalize *multiple (or plural) causation*. Evidence may sometimes show that a given

cause is not the *only* cause (Zelditch, 1971, p. 299). Disjunction or addition indicates that a causal condition can be replaced by another in producing the outcome.

Multiple causation can be expressed in the following way: a given condition C_1 is sufficient to produce the outcome E. However, since it is not the only possible cause, the same outcome E can be produced by another sufficient condition C_2. This is what the disjunction (Boolean addition [+]) signifies. The equation is:

$$E = C_1 + C_2$$

According to Method 2 based on causes to establish sufficient conditions, if C_1 is a sufficient condition, when $C_1 = 1$ we must always find $E = 1$, that is, $P(C_1 \mid E) = 1$ and never $E = 0$, that is, $P(C_1 \mid {\sim}E) = 0$. Similarly, if C_2 is a sufficient condition, when $C_2 = 1$ we must always find $E = 1$, that is, $P(C_2 \mid E) = 1$ and never $E = 0$, that is, $P(C_2 \mid {\sim}E) = 0$. If this is the case, both C_1 and C_2 are sufficient to produce E. However, neither is necessary as Cases 4 and 5 show (the outcome occurs when $C_1 = 0$ or $C_2 = 0$).

Cases	C_1	C_2	$C_1 + C_2$	E
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	0	1	1
5	0	1	1	1

Multiple causation represents disjunctive configurations (OR, +) in which C_1 and C_2 are *sufficient but not necessary* conditions.

Conjunctions and Combinatorial Causation

2. *Necessary but not sufficient conditions*. Conjunctions or Boolean multiplications (AND) are important because they allow to formalize *combinatorial causation*. Data may show that a given cause does not produce an effect on its own but only in conjunction with another. This

means that a causal condition must be accompanied by another to produce the outcome.

Combinatorial causation can be expressed in the following way: a given condition C_1 is necessary to lead to the outcome E. However, since it is not a sufficient condition, the outcome E can only be produced if accompanied by another sufficient condition C_2. This is what the conjunction (Boolean multiplication [•]) signifies. The equation is

$$E = C_1 \cdot C_2$$

According to Method 3 based on effects, if C_1 is a necessary condition for E, when $E = 1$ we must always find $C_1 = 1$, that is, $P(C_1 \mid E) = 1$ and never $C_1 = 0$, that is, $P({\sim}C_1 \mid E) = 0$. Similarly, if C_2 is a necessary condition for E, when $E = 1$ we must always find $C_2 = 1$, that is, $P(C_2 \mid E) = 1$ and never $C_2 = 0$, that is, $P({\sim}C_2 \mid E) = 0$. If this is the case, both C_1 and C_2 are necessary for E. However, none is sufficient on its own as Case 4 (for C_1) and Case 5 (for C_2) show.

Cases	C_1	C_2	$C_1 \cdot C_2$	E
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	0	0	0
5	0	1	0	0

Combinatorial causation represents conjunctive configurations (AND, •) in which C_1 and C_2 are *necessary but not sufficient* conditions.

Combining Connectives

3. *Neither sufficient nor necessary conditions*. Let us take a more complicated example in which no condition is either sufficient nor necessary, but in which two combinations of conditions are both sufficient to produce the outcome E:

$$E = (C_1 \cdot C_2) + (C_3 \cdot {\sim}C_4)$$

If we take the four potential causal conditions separately, neither is sufficient nor necessary as it appears in the following table:

Cases	C_1	C_2	$C_1 \cdot C_2$	C_3	C_4	$C_3 \cdot \sim C_4$	E
1	1	1	1	1	0	1	1
2	1	1	1	1	0	1	1
3	1	1	1	1	0	1	1
4	0	1	0	0	0	0	1
5	1	0	0	1	0	1	1
6	0	1	0	0	1	0	0
7	1	0	0	1	1	0	0

First, it is easy to see that no condition is sufficient on its own to produce the outcome E because all C_i are present when E does not occur (in the last two rows: Cases 6 and 7) according to Method 2. Second, by selecting cases in which $E = 1$ (*Method 3*), that is, when the outcome occurs, it is possible to eliminate all four C_i as necessary conditions, as E also occurs when the potential causal conditions are absent (in Cases 4 and 5). Therefore, none of the four potential causal conditions are either sufficient or necessary for E.

However, the expression above indicates that the combination between C_1 and C_2 produces E or, alternatively, that the combination between C_3 and the absence of C_4 (when $C_4 = 0$) produces E. Both combinations $C_1 \cdot C_2$ and $C_3 \cdot \sim C_4$ are sufficient to produce the outcome: each time the first combination is present the outcome occurs and the same applies to the second combination. However, neither is necessary, as the outcome E occurs also when the two combinations do not occur (Cases 4 and 5).

4. *Necessary and sufficient conditions*. Finally, conjunctions and disjunctions can be used to interpret conditions that are both necessary and sufficient.

- A *sufficient* condition C_1 means that there is no need for it to be in conjunction (through the AND connective) with any other variable C_i to produce E. C_1 is one that by itself always produces the outcome: $P(C_1 \mid E) = 1$ and $P(C_1 \mid \sim E) = 0$.

- A *necessary* condition C_1 means that it cannot be replaced (through the disjunctive OR connective) by any other variable C_i to produce E. C_1 must always be present to produce the outcome: $P(C_1 \mid E) = 1$ and never E = 0, that is, $P(\sim C_1 \mid E) = 0$.

In Boolean terms,

$$E = C_1$$

representing the following table where C_1 is both sufficient and necessary:

Cases	C_1	E
1	1	1
2	1	1
3	1	1
4	1	1
5	1	1

Simplifying Data

As seen, compound statements are based on different combinations of the two connectives conjunction and disjunction ("sums-of-products"), as well as the negation. This allows complex but sometimes long statements. A number of logical devices are useful to simplify data.

1. *Minimization*. The first device for simplifying causal statements is minimization. This tool eliminates causal conditions that appear in one combination of factors (conjunction) but not (disjunction) in another combination of factors otherwise equal to the first one. If *only one* causal condition is different between two combinations of factors both producing the outcome E (say, C_3 is present in one while absent in the other), this causal condition can be considered irrelevant for the outcome.

Take three factors C_1, C_2, and C_3 whose simultaneous presence (through the connective AND) is sufficient to produce E. Imagine further that a second combination of factors (through the connective OR) is also sufficient to produce E. In the second combination, however, C_3 is absent ($\sim C_3$). We have the following compound statement:

$$E = (C_1 \cdot C_2 \cdot C_3) + (C_1 \cdot C_2 \cdot \sim C_3)$$

These are two alternative combinations of factors both sufficient to produce E. None of the single factors C_i are sufficient (see Cases 4 to 6) nor are they necessary (see Cases 7 to 9) on their own for E. However, the two combinations are sufficient conditions (but not necessary). In the first combination three C_i factors are present. Therefore, $C_1 \cdot C_2 \cdot C_3 = 1$. In the second combination the first two C_i are present but C_3 is absent. Thus, $C_1 \cdot C_2 \cdot \sim C_3 = 1$.

Cases	C_1	C_2	C_3	$C_1 \cdot C_2 \cdot C_3$	C_1	C_2	C_3	$C_1 \cdot C_2 \cdot \sim C_3$	E
1	1	1	1	1	1	1	0	1	1
2	1	1	1	1	1	1	0	1	1
3	1	1	1	1	1	1	0	1	1
4	0	1	1	0	0	1	1	0	0
5	1	0	1	0	1	0	1	0	0
6	1	1	0	0	1	1	0	0	0
7	0	1	1	0	0	1	1	0	1
8	1	0	1	0	1	0	1	0	1
9	1	1	0	0	1	1	0	0	1

It is clear that the presence or absence of C_3 is not influential in producing the outcome E and therefore can be eliminated. Whether or not C_3 is present, E occurs anyway. The "*primitive*" *statement* can be simplified into the following "*minimized*" *statement*:

$$E = C_1 \cdot C_2$$

The combination $C_1 \cdot C_2$ is a sufficient condition for E.

The logic on which minimization is based is eminently an experimental one in which between two combinations of factors both producing the same outcome E, *there is only one varying factor* (present in one combination and absent in the other). According to the Method of Difference in an MSSD framework, the only factor that varies while the outcome is constant and all other factors are constant, this can be eliminated as a causal factor.

2. *Implication*. The second device for simplifying causal statements is the implication or the use of "prime implicants." Prime implicants are minimized statements that cover more than one primitive statement. In the example above, the minimized prime implicant $(C_1 \cdot C_2)$ covers both $(C_1 \cdot C_2 \cdot C_3)$ and $(C_1 \cdot C_2 \cdot \sim C_3)$. It is said to imply, cover, or include them. Primitive statements are *subsets* of the prime implicant. Both $(C_1 \cdot C_2 \cdot C_3)$ and $(C_1 \cdot C_2 \cdot \sim C_3)$ are a subset of $(C_1 \cdot C_2)$. The membership of both $(C_1 \cdot C_2 \cdot C_3)$ and $(C_1 \cdot C_2 \cdot \sim C_3)$ is included in the membership of $(C_1 \cdot C_2)$.

In some cases, *several* prime implicants cover the *same* primitive statements. Prime implicants themselves are therefore redundant and minimized statements can be further simplified. This leads to a maximum of parsimony in which only essential prime implicants appear in the causal statement.

Take four cases in which E occurs and we wish to establish which of three potential causal conditions are sufficient and/or necessary. In the following table one sees that there are four alternative combinations (AND) of the C_i producing the outcome E linked through a disjunction OR since all are alternative sufficient conditions for E:

Cases	C_1	C_2	C_3	E
1	1	0	1	1
2	0	1	0	1
3	1	1	0	1
4	1	1	1	1

The primitive statement for this table is the following, where each product corresponds to a row (case) in the previous table:

$$E = (C_1 \cdot \sim C_2 \cdot C_3) + (\sim C_1 \cdot C_2 \cdot \sim C_3) + (C_1 \cdot C_2 \cdot \sim C_3) + (C_1 \cdot C_2 \cdot C_3)$$

According to the minimization principle discussed above

Cases	1 4	$(C_1 \cdot \sim C_2 \cdot C_3)$ $(C_1 \cdot C_2 \cdot C_3)$	minimize into	$(C_1 \cdot C_3)$
Cases	2 3	$(\sim C_1 \cdot C_2 \cdot \sim C_3)$ $(C_1 \cdot C_2 \cdot \sim C_3)$	minimize into	$(C_2 \cdot \sim C_3)$
Cases	3 4	$(C_1 \cdot C_2 \cdot \sim C_3)$ $(C_1 \cdot C_2 \cdot C_3)$	minimize into	$(C_1 \cdot C_2)$

The minimized statement is therefore

$$E = (C_1 \cdot C_3) + (C_2 \cdot \sim C_3) + (C_1 \cdot C_2)$$

These three prime implicants, however, cover the following primitive statements:

$(C_1 \cdot C_3)$	covers	$(C_1 \cdot \sim C_2 \cdot C_3)$ $(C_1 \cdot C_2 \cdot C_3)$	
$(C_2 \cdot \sim C_3)$	covers	$(\sim C_1 \cdot C_2 \cdot \sim C_3)$ $(C_1 \cdot C_2 \cdot \sim C_3)$	
$(C_1 \cdot C_2)$	covers	$(C_1 \cdot C_2 \cdot \sim C_3)$ $(C_1 \cdot C_2 \cdot C_3)$	(already covered by $[C_2 \cdot \sim C_3]$) (already covered by $[C_1 \cdot C_3]$)

Therefore, $(C_1 \cdot C_2)$ is a redundant prime implicant and can be eliminated

$$E = (C_1 \cdot C_3) + (C_2 \cdot \sim C_3)$$

meaning that E is caused either by the multiplication $(C_1 \cdot C_3)$ or by the multiplication $(C_2 \cdot \sim C_3)$. Both are sufficient but not necessary conditions for E (as each can be replaced by the other combination).

3. *Factorization*. The third device for simplifying causal statements is the factorization. More precisely, factorization helps in *clarifying* the structure of the data rather than simplifying it.

First, factorization helps highlighting *necessary conditions*. In the following causal statement:

$$E = (C_1 \cdot C_3) + (C_2 \cdot C_3)$$

C_3 is a necessary (but not sufficient, see Case 6) condition, whereas C_1 and C_2 are neither necessary (Cases 7 and 8) nor sufficient (Cases 4 and 5), as it appears in the table below. On the contrary, the two alternative (OR) combinations (AND) $C_1 \cdot C_3$ and $C_2 \cdot C_3$ are sufficient conditions for E.

Cases	C_1	C_2	C_3	$C_1 \cdot C_3$	$C_2 \cdot C_3$	E
1	1	1	1	1	1	1
2	1	1	1	1	1	1
3	1	1	1	1	1	1
4	1	0	0	0	0	0
5	0	1	0	0	0	0
6	0	0	1	0	0	0
7	0	1	1	0	1	1
8	1	0	1	1	0	1

Factoring the causal statement above is useful to show C_3 as a necessary condition:

$$E = C_3 \cdot (C_1 + C_2)$$

Second, factorization helps identifying *causally equivalent sufficient conditions*. In this example, C_1 and C_2 are equivalent in their combination with C_3 to produce two different combinations both of which are sufficient conditions for E. It does not matter (it is equivalent) with which condition C_3 combines. In both combinations it produces a sufficient condition.

Beyond Dichotomization: Fuzzy Sets and the Use of Computer Programs

Because Boolean logic is a form of algebra in which all values are reduced to either "true" or "false," it has been crucial to the development of computer science based on 0/1 bit systems. Quite naturally, therefore, computer programs for the analysis of necessary and sufficient conditions with dichotomous data have developed in a number of fields, in particular in the fields of linguistics and text information retrieval (Zadeh, 1965), and search engines on the Internet. Initially based on dichotomous 0/1 systems, these retrieval methods have evolved to consider the frequency of terms in documents, allowing to weight information and transform systems to include *ordinal* or "*fuzzy*" *data* (Kraft, Bordogna, & Pasi, 1994; Meadow, 1992).

Social sciences have followed this evolution with Ragin's pioneering work on computer programs for Boolean analysis. The computer program for dichotomous data that was developed by Ragin and his collaborators—*Qualitative Comparative Analysis* (QCA)—was inspired by algorithms created by electrical engineers in the 1950s (Drass & Ragin, 1986, 1992; McDermott, 1985). Recently, a new software has been developed (Ragin, Drass, & Davey, 2003; Ragin & Giesel, 2003) to include fuzzy sets (*Fuzzy Set/Qualitative Comparative Analysis* or FS/QCA). Both programs are widely used and give rise to an increasing number of studies.[22]

The principles and rules for establishing sufficient and necessary conditions do not change when moving from dichotomous variables (0/1) to *ordinal* and *interval* (or ratio) variables. The aim of the comparative method is to identify sufficient and/or necessary conditions in the form of single attributes or, more typically, in the form of *configurations* (through specific combinations of attributes). Whether such configurations are constructed from 0/1 variables or from ordinal variables does not change the method of assessing if they are sufficient or necessary conditions for an outcome to occur.

Take the variable "state formation" operationalized as follows: "before 1815 (1)," "between 1815 and 1914 (2)," and "after World War I (3)." This variable can combine in different ways with a similar operationalization of "industrialization" "before 1870 (1)," "between 1870 and 1914 (2)," and "after World War I (3)." There are nine different configurations possible. With the methods above we can test which is either a necessary or sufficient condition (or both) for, say, high levels of national integration. Such combinations can also be done with interval or ratio variables: literacy rates among the adult population or urban density levels.

A recent way to move beyond dichotomization is to use *fuzzy-set approaches* (Mahoney, 2000, 2003; Ragin, 2000). Instead of being based on conventional "crisp" sets in which a case is either "in" or "out" (0/1) as in classical categorization, fuzzy sets allow membership in the interval between 0 and 1. For example, in a crisp set a family may be either "financially secure" or not. In a fuzzy set a family may be "almost" financially secure, say .85, that is, part of the set financially secure but not completely. *Fuzzy membership scores* are given to cases according to their degree of membership to the set. The United States does not fully belong to the set of "democracies" but almost (with a value of .80 according to Ragin, 2000, p. 176). This follows alternative ways of categorizing data such as family resemblance and radial categories.

Whereas in "variable-oriented" research categories are created *from* the values of the cases (a safe neighborhood is one in which the crime rate is, say, below 5% and a financially secure family is one with an income above, say, $40,000), in fuzzy sets "measurement" is carried out by attributing values to cases on the basis of the degree of belonging to a category or set. This is simply a different way of attributing *values to cases according to a given property* in which the knowledge of specific cases by the researcher plays a bigger role.

Finally, as with 0/1 data, compound statements are again formulated with the aid of *operators or connectives*—NOT, AND, and OR being the most important ones. With respect to conventional sets there are, however, some differences (Ragin, 2000, pp. 171–178). The two following subsections deal with the differences in the formulation of conditional statements. A discussion of how values are attributed to cases is not included.

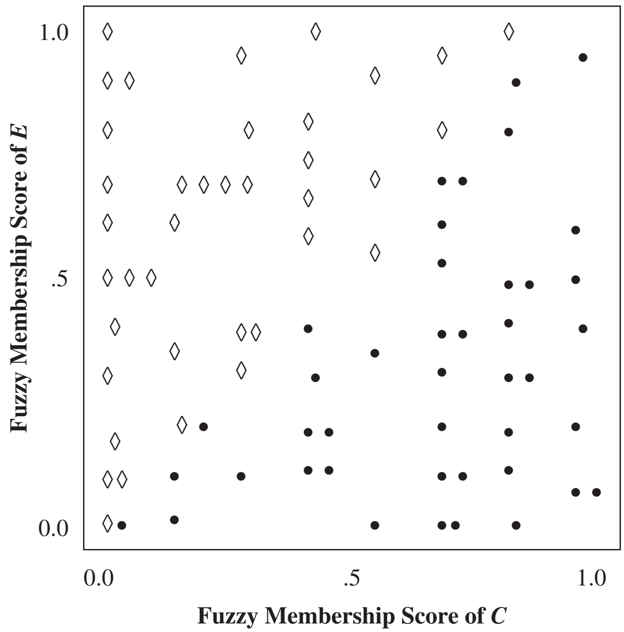
Necessary and Sufficient Conditions

If C is a necessary condition for an outcome E, for all instances in which E is present, C must be present too. If "losing a war" is necessary for "social revolution," then there must be no social revolutions without "losing a war:" $P(\sim C \mid E) = 0$. However, we may find cases of "losing a war" where no social revolution has taken place ("losing a war" is not a sufficient condition). Thus, if all instances of $E = 1$ must also have $C = 1$ but there might be instances of $C = 1$ in which $E = 0$, then $E = 1$ is a subset of $C = 1$. Imagine 15 countries among which 10 have a PR electoral system and that, among these, 8 are MPS and 2 are TPS. All MPS have PR ($P(\sim C \mid E) = 0$) but not all PR lead to MPS (there are two "exceptions"). MPS = 1 (dependent variable) is a subset of PR = 1 (independent variable).

If the discrete values 0 and 1 are replaced by fuzzy values between the two extremes (degrees of proportionality of electoral systems and effective number of parties), this logic does not change. If a higher degree of proportionality is a necessary condition for a higher number of parties in a party system, then we must not find cases of higher number of parties with low levels of proportionality: $P(\sim C \mid E) = 0$. On the other hand, all instances of high number of parties must have also a high level of proportionality: $P(C \mid E) = 1$. However, we can have a high proportionality but few parties (since proportionality is necessary but not sufficient). As before, instances of many parties are a subset of instances with high proportionality.

The scattergram below depicts the theoretical distribution of values if C (proportionality) is a necessary condition for many parties (symbolized with •). When researchers find instances in which scores in the outcome are *less than* (*or equal to*) scores in the cause, then it is possible to conclude that we are in the presence of a necessary condition.

If C is a sufficient condition for an outcome E, for all instances in which C is present, E must be present too. If "losing a war" is sufficient for "social revolution," then there must be no "losing a war" without "social revolutions:" $P(C \mid \sim E) = 0$. However, we may find cases of social revolution that did not



lose a war ("losing a war" is not a necessary condition: the same *E* can be caused by another factor such as "repressive regime"). Thus, if all instances of $C = 1$ must also have $E = 1$ but there might be instances of $E = 1$ in which $C = 0$, then $C = 1$ is a subset of $E = 1$. Imagine 15 countries among which 10 have MPS and that, among these, 8 are ethnically fragmented (FRAG) and 2 ethnically homogeneous. All FRAG are followed by MPS ($P(C \mid E) = 1$) but not all MPS need FRAG (there are two "exceptions"). FRAG = 1 (independent variable) is a subset of MPS = 1 (dependent variable).

Replacing discrete values with fuzzy values, if a higher degree of ethnic fragmentation is a sufficient condition for many parties, then we must not find cases of high ethnic diversity with few parties: $P(C \mid {\sim}E) = 0$. However, we can have a high number of parties without ethnic fragmentation (since fragmentation is not necessary and can be replaced by another factor such as PR). As before, instances of ethnic fragmentation are a subset of instances with many parties.

The theoretical distribution of values if *C* (ethnic fragmentation) is a sufficient condition for many parties is symbolized with ◊ in the scattergram above. When researchers find instances in which scores in the outcome are *more than* (*or equal to*) scores in the cause, then it is possible to conclude that we are in the presence of a sufficient condition.

Compound Statements

As above, the discussion is limited to the three main operators NOT, AND, OR.

1. *Negation: NOT*. In datasets with dichotomous variables the negation is the contrary of the value: the negation of 0 is 1 and vice versa. In fuzzy datasets the negation is given by the subtraction of the fuzzy membership score from 1:

Negation of fuzzy membership in set A = 1 − [fuzzy membership score in set A]

For example, if the fuzzy membership score of Britain in the set "PR electoral systems" is .10, its negation (i.e., fuzzy membership in the set "non-PR systems") is .90:

$$\sim.10 = 1 - .10 = .90$$

In the table below the column $\sim C_1$ gives examples of negation scores for C_1.

2. *Conjunction: AND*. In datasets with dichotomous values the conjunction occurs when several factors must be present to produce an outcome ($C_1 \cdot C_2$). Both factors must have a value of 1 to produce the outcome. In fuzzy data sets, cases may have different degrees of membership in the sets represented by C_1 and C_2 (see again the table below for an example). The fuzzy membership score of a case in the "conjunction set" of both factors is established by taking the *minimum* membership score.

Consider again a statement about the causes of MPS. In a dichotomous dataset, a hypothesis could be that the combination of PR and FRAG is a sufficient condition for producing MPS: PR · FRAG = MPS. If both are present, MPS is also present. To establish if a country is a "member" of the set of countries that have both PR and FRAG, we take the *minimum*.

Cases	*PR* C_1	*FRAG* C_2	*Negation* $\sim C_1$	*Conjunction* $C_1 \cdot C_2$	*Disjunction* $C_1 + C_2$
Britain	.10	.40	.90	.10	.40
Belgium	.95	.80	.05	.80	.95
Italy	.40	.20	.60	.20	.40

If a country, for example, the United States or India in the table below, score 0 on PR and 1 on ethnic fragmentation, the value of the compound statement PR · FRAG = 0, that is, the *smaller* value between PR = 0 and FRAG = 1. The same applies if we replace discrete values with fuzzy values. Suppose a country, say Britain in the table above, for which the fuzzy membership score on PR (the set of PR electoral systems) is .10 and on FRAG (the set of ethnically fragmented countries) is .40. In this case, the membership in the set of countries that are both proportional and fragmented is .10.

Cases	*PR* C_1	*FRAG* C_2	*Negation* $\sim C_1$	*Conjunction* $C_1 \cdot C_2$	*Disjunction* $C_1 + C_2$
United States	.00	1.00	1.00	.00	1.00
India	.00	1.00	1.00	.00	1.00

3. *Disjunction: OR*. The disjunction OR is the other most common operator used for compound statements. In conventional data sets, the disjunction occurs when one or another factor is present to produce an outcome ($C_1 + C_2$). At least one of the two factors must have a value of 1 to produce the outcome, but not necessarily the two. In fuzzy data sets, cases may have different degrees of membership in the sets represented by C_1 and C_2. Contrary to the conjunction, the fuzzy membership score of a case in the "disjunction set" of several factors is established by taking the *maximum* membership score.

Taking again the same example, we may formulate the hypothesis that the compound statement PR or multimember constituencies (MM) is a necessary condition for MPS to occur, that is, that either PR or MM must be present but not necessarily both (PR + MM = MPS). A large constituency magnitude may have the same "proportionalizing" effects than PR even if the electoral formula is majoritarian. However, if both are absent the effect is not produced.

If a country, for example, Britain in the 19th century when most constituencies were multimember, score 0 on PR and 1 on magnitude, the value of the compound statement (PR + MM) is 1, that is, the *larger* value between PR = 0 and MM = 1. The same applies if we replace discrete values with fuzzy values. Suppose a country, say again Britain, for which the fuzzy membership score on PR (the set of proportional electoral systems) is .10 and on MM (the set of countries with multimember constituencies) is .70. In this case, the membership in the set of countries that have either PR or multimember constituencies is .70.

With these operators it is possible to formulate compound causal statements in terms of necessary and sufficient conditions as described in the previous subsection. These techniques—especially when powered by computerized software—allow sophisticated analyses well beyond the basic principles presented here.