4

The Uses of "Negative" Cases in Social Research

This chapter examines three approaches to the analysis of dichotomous outcomes: conventional quantitative analysis, qualitative comparative analysis (QCA), and analytic induction (AI).[1] My goal is to highlight the distinctive features of AI by contrasting it with the other two approaches. The specific focus is on their contrasting uses of "negative" cases. Here, I refer to instances of the presence of an outcome (e.g., employed) as *positive* cases, and to instances of the opposing category (e.g., not employed) as *negative* cases. This usage of *positive* versus *negative* cases should not be confused with an alternate convention, which is to use positive versus negative to differentiate cases that are theory-confirming from those that are theory-disconfirming (Katz 1983; Athens 2006).

Table 4-1 illustrates the difference between positive/negative and confirming/disconfirming, using a 2 × 2 table cross-tabulating the presence/absence of an outcome against the presence/absence of a cause. Cases in cell *b* (cause present/outcome present) are positive and confirming, whereas cases in cell *c* (cause absent/outcome absent) are negative and confirming. Cases in cell *a* (cause absent/outcome present) are positive but disconfirming, whereas cases in cell *d* (cause present/outcome absent) are both negative and disconfirming.

The three approaches to dichotomous outcomes addressed in this chapter can be arrayed along a continuum with respect to the dependence of standard applications of each approach on the analytic incorporation of "negative" cases. Conventional quantitative analysis is fully dependent on negative cases, and its treatment of negative cases is fully symmetrical with its treatment of positive cases. Without variation in the dependent variable (i.e., without both positive and negative cases of a dichotomous outcome), there is nothing to explain. Most applications of the second approach, QCA, are also dependent on negative cases, but in a different manner. QCA's truth table procedure uses negative cases to classify truth table rows as true or false based on the degree to which the cases in each row

TABLE 4-1 Simple cross-tabulation of a causal condition and an outcome

	Cause absent	Cause present
Outcome present	*a* = positive and disconfirming	*b* = positive and confirming
Outcome absent	*c* = negative and confirming	*d* = negative and disconfirming

consistently display a given outcome. As explained in this chapter, because the truth table approach focuses on the *consistency* of the link between causal conditions and *positive* outcomes, it is best understood as "partially asymmetric." Finally, negative cases of the outcome play no direct role in AI, which separates the analysis of positive cases from the analysis of negative cases, basically eschewing the concept of negative cases altogether. In this "fully asymmetric" approach, negative cases are viewed as positive cases of one or more alternate outcomes.

An important first step in this discussion is to recognize that most dichotomies in the social sciences are not empirically binary (Goertz and Mahoney 2012: 161–65).[2] One side of the dichotomy—usually the focal category—is well defined and relatively homogeneous, while the other side, the "complement," is typically heterogeneous, with cases united only by their non-membership in the named side of the dichotomy.[3] For example, a researcher might be interested in the difference between voting Republican (the focal category—positive cases) and not voting Republican (the opposing category—negative cases), without differentiating among the different kinds of negative cases included in the complement of the focal category (e.g., voting Democratic, voting for a third party, refusing to vote, forgetting to vote, or deliberately casting an invalid ballot, to name a few).[4] One of the main points of this chapter is that AI addresses each outcome separately and rejects treating heterogeneous complements (as in "not voting Republican") as if they are homogeneous. This view of negative cases contrasts sharply with conventional practices in both quantitative research and most applications of QCA, where membership in the focal category versus membership in its heterogeneous complement is often the main focus of the analysis, and cases included in a heterogeneous complement are rarely differentiated according to the alternate outcomes they display.[5] In fact, a central conclusion of the discussion that follows is that AI challenges the very notion of "negative" cases, even in situations where the outcome in question is empirically binary.

NEGATIVE CASES IN CONVENTIONAL QUANTITATIVE RESEARCH

The simplest variable type in conventional quantitative research is the dichotomy. Dichotomies are often used to signal the presence/absence of some trait or outcome (e.g., married vs. not married) and are typically dummy-coded, with 1 = present or *yes* and 0 = absent or *no*. The assignment of 1 or 0 to categories is completely arbitrary; it is determined by the researcher according to which side of the dichotomy makes more sense as the reference category (which is then coded 0 on the dummy variable).[6]

TABLE 4-2 Hypothetical cross-tabulation of "Married" and "Voted Republican"

	Not married (0)	Married (1)
Voted Republican (1)	$a = 300$	$b = 250$
Did not vote Republican (0)	$c = 500$	$d = 30$

In conventional quantitative research, dichotomies are treated as though they are fully symmetrical, which is consistent with the arbitrariness of their 1/0 coding. Their symmetrical nature is apparent in analyses of their relations with other variables. Consider, for example, table 4-2, which shows a hypothetical cross-tabulation of married versus not married (conceived as an independent variable) against voted Republican versus its complement, did not vote Republican (conceived as a dependent variable).

Because conventional quantitative analysis is fully symmetrical, cases in cells *b* and *c* count in favor of a relation between being married and voting Republican, equally so, while cases in cells *a* and *d* count against this argument, again equally so. Expressed in log-odds terms, the connection between being married and voting Republican is

$$\text{log odds Republican} = -0.05108 + 2.6311 \bullet (\text{married}) + \text{e}$$

Reversing the 1/0 coding of the dependent variable, the equation for the effect of married on the log odds of not voting Republican is

$$\text{log odds of not Republican} = 0.05108 - 2.6311 \bullet (\text{married}) + \text{e}$$

In short, the same exact absolute coefficients are attached to the constant and the slope; only the signs are reversed. It is thus reasonable to refer to the complement (the negated pole) in conventional uses of dichotomous outcomes as being "fully symmetrical" with the focal category. The focal category and its complement are analytically equivalent and mathematically interchangeable. Of course, this feature of complements is well known to quantitative researchers.

It is important to point out that quantitative analysis of a dichotomous outcome focuses directly on differences between the focal outcome and its complement. Conventional quantitative analysis without variation is impossible, and the focal outcome and its complement must be analytically paired. They are mutually constitutive and, in a sense, "codependent."

NEGATIVE CASES IN QCA

QCA is grounded in the analysis of set relations and truth tables. Negative cases come into play in two major ways: (1) they are used in the assessment of the consistency of the degree to which cases sharing one or more causal conditions agree in displaying a given outcome; and (2) they impact the assignment of outcome codes to truth table rows, which summarize the different combinations of conditions linked to an outcome (see appendix A). I discuss these two uses of negative cases in turn, limiting the discussion to crisp sets in order to simplify the presentation. The extension to fuzzy sets is straightforward (see Ragin 2000, 2008; Ragin and Fiss 2017; appendix B).

QCA partitions cross-tabulations like the one in table 4-2 into different set relations, depending on the focus of the investigation (Ragin 2008: 13–28). For example, in set-analytic research it is common to assess the degree to which cases that display a causal condition (e.g., married) constitute a more or less consistent subset of the cases displaying the outcome (e.g., voted Republican). If the proportion (p) of consistent cases (cell *b* divided by the sum of cells *b* and *d)* is very high (e.g., $p \geq 0.85$), then the researcher may conclude that the causal condition (married) is usually sufficient for the outcome (voted Republican). A less controversial way of stating this connection is simply to observe that the outcome (voted Republican) is "widely shared" by cases with the causal condition in question (married). This set-theoretic relation focuses exclusively on cases in the second column of table 4-2. Thus, the calculation of the degree to which a causal condition is a *consistent* subset of the outcome uses the negative cases residing in cell *d*, but not those in cell *c*.

Another key set-theoretic relation is the degree to which instances of the outcome constitute a subset of instances of a causal condition—or, more simply, the degree to which instances of the outcome share a given antecedent condition. When an outcome is a subset of a causal condition, the interpretation of the causal condition as necessary but not sufficient may be warranted (Braumoeller and Goertz 2000; Dion 1998; Goertz 2020, 2017). This set-analytic assessment focuses exclusively on the first row of table 4-2; the proportion of cases consistent with this set relation is the number of cases in cell *b* divided by the sum of cases in cells *a* and *b*. If cell *a* is empty and cell *b* is well populated with cases, then the evidence is fully consistent with the set-theoretic relation in question. Note, however, that this calculation does not involve negative cases, but instead focuses exclusively on cases displaying the outcome.

Neither of the two assessments central to the set-theoretic analysis of table 4-2 involves the negative cases in cell *c*, the "null-null" cell (e.g., not married/did not vote Republican). Thus, a cell that is central to conventional quantitative analysis—cases in this cell count in favor of the researcher's argument that the two variables are correlated—is not directly relevant to either of the two main set-analytic assessments of table 4-2. Because cases in cell *c* have no direct relevance to the two central assessments in the set-theoretic approach, the approach can be described as "partially asymmetric" in its consideration of three of the four cells of the table. Cases in cell *c* become relevant to the set-analytic approach only if the researcher in this example shifts attention from the analysis of voting Republican to the analysis of not voting Republican.[7]

Assessing the consistency of the degree to which cases that share one or more causal conditions agree in displaying a given outcome is central to truth table analysis, a core QCA procedure. Truth tables list the logically possible combinations of causal conditions and assign an outcome code to each combination. Outcome codes can be true (1), false (0), or undetermined (?), based on both the number of cases assignable to each truth table row and the consistency of the set relation. Essentially, the consistency score for each row assesses the degree to which membership in the row is a subset of membership in the outcome. In other words, these scores assess the degree to which cases in each truth table row agree in displaying the outcome, using (for crisp sets) the number of cases in each row displaying the outcome divided by the total number of cases in each row:[8]

$$(n \text{ positive cases}) \;/\; (n \text{ positive cases} + n \text{ negative cases})$$

Of course, truth table rows vary in their degree of consistency, and the researcher must select a threshold value (e.g., $p \geq 0.85$) for a truth table outcome code of *true* (a value of 1). The important point is that negative cases have a huge impact on truth table analysis via their role in the calculation of subset consistency scores, which in turn determine the outcome coding of truth table rows.

Note that QCA's set-analytic approach to negative cases shares an important feature with the conventional quantitative approach. Specifically, the complement of the focal outcome category is treated as just another category. There is no allowance for the fact that the set of negative cases may be heterogeneous and therefore may constitute a set that is qualitatively different from the focal category (i.e., the set of positive cases).

NEGATIVE CASES AND ANALYTIC INDUCTION

AI offers a different template for the treatment of set complements. Its distinctive approach to complements stems in large part from its affinity for "How did it happen?" questions in social research (see chapter 3). Howard Becker (1998: 196) states that AI "is ideally suited to answering 'How?' questions, as in 'How do these people do *X*?'" How does one become a marijuana user (Becker 1953), an opiate addict (Lindesmith 1968), or an embezzler (Cressey 1973)? How does collective violence erupt? What about military coups? Questions like these place positive instances of outcomes front and center.

AI seeks to identify relevant antecedent conditions shared by positive instances. Using table 4-2 terminology, the goal is to establish that cell *a* is empty, while cell *b* is well populated with cases. Thus, the primary focus is on the first row of table 4-2, which overlaps with one of the major concerns of QCA's set-analytic approach. Using AI, however, disconfirming cases in cell *a* are treated as prods to further research, which may lead, in turn, to a conceptual realignment of the evidence, as discussed in detail in chapter 2. The strategic goal is to increase the consistency of the connection between causal conditions and the outcome, removing cases from cell *a* by eliminating them from the analysis altogether (e.g., via scope conditions) or by moving them from cell *a* to cell *b* or *c* via some form of conceptual realignment.

While AI's primary focus is on cases in cells *a* and *b*, it is important to address AI's approach to cases in cell *d* as well.[9] After all, cases in cell *d*—instances of the causal condition (or combination of conditions) that nevertheless failed to display the outcome—are essential foot soldiers in Robinson's (1951) broadside against AI (see chapter 1). Recall that AI's goal is to answer "How did it happen?" questions, and that negative cases (i.e., plausible candidates for the outcome that nevertheless did not experience it) are not directly relevant to this task. With regard to negative cases, however, AI asks, "What happened instead?" Despite experiencing favorable antecedent conditions, cell *d* cases did not experience the focal outcome. The AI researcher's task is to examine these cases and identify the varied, alternate outcomes they experienced, thereby specifying the heterogeneity of the complement of the focal outcome. For example, while the focal category "voted Republican" is relatively uniform and well circumscribed, there are several different ways for people to attain membership in the complement, "did not vote Republican"—not voting, voting Democratic, voting for a third party, deliberately casting an invalid ballot, and so on. These alternate outcomes should be studied separately with an eye toward the antecedent conditions specific to each. That is, each alternate outcome may be deserving of separate consideration as positive instances of something else.

Recognizing the diversity of negative cases can be a first step toward developing a typology of outcomes (George and Bennett 2005). For example, Theda Skocpol's (1979) study *States and Social Revolutions* discusses several cases that did not culminate in revolution. Rather than treating them simply as instances of "not revolution," she categorizes them in terms of what happened instead. England had a political revolution rather than a social revolution; Japan had a revolution from above rather than a social revolution; Germany had a successful revolt that did not culminate in revolution; and so on.

AI's proper response to Robinson's (1951) critique is as follows: (1) yes, if cases exhibiting the relevant causal conditions but not the focal outcome exist, they do matter; (2) such cases are usually heterogeneous and should be differentiated according to their separate outcomes; (3) these alternate outcomes should be viewed as happenings in their own right; and (4) each alternate outcome may be subjected to the same type of analytic scrutiny that instances of the focal outcome receive (Kidder 1981). In short, so-called negative cases should be understood as positive instances of other, alternate outcomes.

Consider the following example. A researcher interested in cases of electoral fraud in developing countries identifies a set of countries in which national elections are either scheduled or planned, and follows them over time. Electoral fraud occurs in a substantial number of these countries. The researcher completes an application of AI and identifies four antecedent conditions shared by positive cases of electoral fraud: unpopular regime, clientelistic political system, chief executive who dominates the military, and a viable opposition party or coalition. Using the

TABLE 4-3 Hypothetical study of electoral fraud

	One or more of the four antecedent conditions absent	Antecedent conditions present: unpopular regime, clientelism, executive dominates military, vigorous opposition party
Electoral fraud	Cell *a:* no cases here	Cell *b:* electoral fraud cases here
No electoral fraud	Cell *c:* cases lacking electoral fraud and one or more antecedent conditions	Cell *d:* cases lacking electoral fraud but displaying the four antecedent conditions; the AI researcher addresses the question "What happened instead?"

language of table 4-1, the researcher's cell *a* is empty, while cell *b* is populated with positive instances of electoral fraud, as illustrated in table 4-3. Further, the researcher certifies that the causal recipe identified via the application of AI resonates with case-level knowledge—that is, it rings true as an account of the conditions linked to electoral fraud in developing countries.

However, the researcher also identifies a substantial number of candidate cases that did not display electoral fraud. Regarding these cases (especially those residing in cell *d*) the researcher asks, "What happened instead?" Suppose the researcher investigates this question for each negative case and identifies three alternate outcomes: (1) instances of regime change prompted by popular uprisings, (2) instances of potential voting fraud that were thwarted by international supervision of elections, and (3) instances of canceled elections amid the imposition of martial law. The researcher decides to push the investigation forward by applying AI to the cases of regime change, with an eye toward conditions that may have prompted or enabled popular uprisings.

While the cases in cell *d* share the four antecedent conditions exhibited by the positive instances of electoral fraud, there is, of course, no guarantee that these four conditions are all relevant as antecedent conditions for the alternate outcome, regime change. In the end, only the conditions that resonate with case-level analysis would be retained as antecedent conditions in an investigation of the subset of cases exhibiting regime change.

A final issue regarding cases in cell *d* is the situation where one of the alternate outcomes is the successful conduct of fair elections (without requiring international supervision). The existence of such cases would seem to validate Robinson's (1951) concerns regarding the limitations of AI: despite sharing the four antecedent conditions experienced by the cases of electoral fraud (cell *b* cases), a subset of the cell *d* cases successfully conducted fair elections. It is important to consider, however, that AI treats alternate outcomes as worthy of separate consideration and analysis—as positive outcomes in their own right. In the course of doing so, it is very likely that the researcher would identify decisive differences between these cases and cell *b* cases. The conditions linked to fair elections in the presence of such adverse circumstances would certainly warrant scholarly attention.

The important point is that AI addresses negative cases in a way that respects their status as alternate outcomes. They are not treated as residual cases, nor are they treated collectively as just another category (i.e., as an undifferentiated set complement). Instead, their diverse outcomes are distinguished and then assessed separately. In this respect, it is clear that AI eschews the concept of negative cases altogether. Negative cases are more properly viewed as positive cases of something else, as alternate happenings.

CONCLUSION

Conventional quantitative analysis uses all four cells in table 4-2 to derive a symmetrical assessment of association, giving all four cells equal voice in the calculation of the nature and strength of the connection between antecedent conditions and outcomes. Likewise, QCA uses negative cases in cell *d* to assess the degree to which cases with different combinations of antecedent conditions share a given outcome, which in turn is the basis for coding truth table rows as true or false. From the viewpoint of AI, the quantitative approach and QCA's set-analytic approach to complements share two important liabilities. In both approaches the focal categories are clearly specified and relatively homogeneous, while the complements are unspecified and potentially heterogeneous. The unspecified nature of set complements is typically ignored in both QCA and conventional quantitative research. The second liability is that negative cases are given a major voice in shaping the researcher's findings regarding the conditions linked to positive outcomes. While this practice may seem perfectly appropriate when the goal is to explain variation in an outcome, it is less so when the goal is to explain how an outcome happens. AI, by contrast, rejects the idea of an unspecified, heterogeneous complement, asking "What happened instead?" and treating alternate outcomes as positive instances of something else.