# 7

# Generalized Analytic Induction

## *A Step-by-Step Guide*

The simplest application of generalized AI is to a set of cases included in an investigation because they all display the same outcome. There are no "negative" cases, per se, and thus no variation or outcome difference to "explain." In the language of conventional quantitative analysis, the outcome is not a variable; rather, it is more or less constant across the cases included in the study. As noted previously, conventional quantitative analysis requires a dependent *variable*; constants are off-limits. Likewise, the parsimonious and intermediate solutions of qualitative comparative analysis require both positive and negative cases, so that "remainder" rows can be defined and manipulated; lacking negative cases, researchers using QCA are able to derive only the complex solution (see chapter 6).

Qualitative researchers often find the definition or circumscription of relevant negative cases problematic. For example, consider a researcher interested in how Olympic athletes sustain their commitment to being Olympic caliber. Defining positive cases is relatively straightforward: the researcher would identify current Olympic athletes who have maintained their commitment for a substantial period. But what are good negative cases and how might they be useful? Nonathletes are clearly irrelevant, as are athletes who are not Olympic caliber. The challenge would be to select an appropriate subset of Olympic athletes who somehow failed to sustain commitment. Perhaps the best negative cases in a qualitative study would be Olympic-caliber athletes who were once clearly committed but failed to sustain their commitment for an extended period.[1]

Note, however, that the conditions that lead to failure to sustain commitment (chronic injury, financial stress, and so on) are likely to be different from (and probably not the simple reverse of) the conditions that sustain commitment (e.g., involvement in a social network of like-minded athletes). While it might be important to know that chronic injury poses an obstacle to the accomplishment of sustained commitment, the primary focus of the investigation in question is on *how* sustained commitment is accomplished, not on factors that pose obstacles to commitment. Instances of the failure to accomplish sustained commitment can provide only limited information about how it is sustained. From the perspective of AI, each outcome is deserving of separate consideration and treatment (see chapter 4). The accomplishment of sustained commitment and the failure to sustain commitment are different outcomes, ruled by different mechanisms. Of course, knowledge of both outcomes would be useful, and the two analyses would undoubtedly complement and inform each other. The important point is that AI separates them.

Using hypothetical data on Olympic-caliber athletes, this chapter offers an example of the application of generalized AI to an analysis of a set of cases displaying the same outcome: sustained commitment. The example also demonstrates how fsQCA software (Ragin 2021; Ragin and Davey 2021) can be used to implement generalized AI.[2]

## GENERALIZED AI: BASIC STEPS

1. Define the outcome of interest. The outcome should be conceived as a qualitative change, for example as a "happening," an "instance," or something that is "accomplished." The outcome can be at any level of analysis (e.g., micro-, meso-, or macro-level; Katz 2001). Also, its precise definition and operationalization should be open to strategic revision as the research progresses, as explained in chapter 2.
2. Identify relevant instances of the outcome. It is more important to have diverse cases that exhibit the outcome than it is to have a strictly representative sample of cases (Goertz and Mahoney 2012). It is also important that the cases selected for analysis are meaningfully related in some way—for example, they could be situated in a specific time and place. The important point is that cases of the outcome should be drawn from a well-defined and circumscribed set.
3. Conduct case-level research in order to identify the central contributing conditions for each case. Remember, the goal is to explain "how" the outcome in question comes about. This research should be guided by theory, but it is important for there to be an inductive aspect as well. If it is not possible to examine all the cases, focus on a diverse subset of cases. Identify the most common contributing conditions.
4. Once a satisfactory set of contributing conditions has been identified, assess the membership of each case in each condition. This step can be either an assessment of the presence/absence of each condition or a fuzzy-set assessment of the degree to which each contributing condition is present.[3]

5. Construct a data spreadsheet describing the cases with respect to the contributing conditions identified in each case. The cases define the rows of the data spreadsheet; the contributing conditions define the columns. Each data cell is either a presence/absence coding of the contributing condition or a fuzzy-set coding of the degree to which the contributing condition is present.
6. Code an outcome value for each case. If the outcome is crisp, code each case with an outcome of 1. If the outcome is a fuzzy set, the cutoff value should be ≥0.5. The dialogue box for setting up the truth table analysis permits the specification of a threshold value when the outcome varies by level or degree. Enter the data into fsQCA's data spreadsheet or transfer the data to fsQCA as a comma delimited file (*.csv) from Microsoft Excel. An example using hypothetical data on twenty Olympic athletes is presented in table 7-1.
7. Using fsQCA, convert the data spreadsheet into a truth table. In the dialogue box that governs the construction of the truth table, the user can specify which side (positive or negative) of each contributing causal condition is expected to be linked to the outcome. Code relevant conditions so that they reflect the interpretive logic of "contributing versus irrelevant" instead of "present versus absent." If a condition is thought to be contributing to the outcome when equal to one (present), the zeros in the condition's truth table column are recoded to dashes, indicating irrelevance. If a condition is thought to be contributing when equal to zero (absent or negated), the ones in the condition's truth table column are recoded to dashes, indicating irrelevance. The researcher has the option of using both conventional presence/absence conditions and contributing/irrelevant conditions.
8. Establish a frequency threshold to filter out low-frequency truth table rows. The goal of the truth table analysis is to identify "modal configurations"—combinations of antecedent conditions that occur with substantial regularity. Usually, a higher frequency threshold will result in modal configurations with more conditions; often, a lower frequency threshold will yield simpler configurations.
9. Run the truth table minimization procedure in order to derive the key combinations of antecedent conditions linked to the outcome. In effect, with this setup, truth table minimization is roughly the same as applying the set "inclusion" rule to the evidence (see examples in chapter 6).
10. Manipulate the resulting equation algebraically to clarify the causal recipes (see chapter 6). For example, check for conditions that can be joined by logical *or* to create a close connection with the outcome. If there are multiple recipes, consider specifying outcome subtypes, following the illustration in chapter 6.

11. Evaluate the results with reference to cases. Are the results consistent with what is known about cases? Do the results resonate with or enrich case-level knowledge? Identify cases that exemplify the causal recipe(s).

## APPLICATION OF GENERALIZED AI

In this example, the researcher studies how twenty Olympic athletes maintain their commitment and finds five widely shared conditions, thus completing steps 1–4 sketched above. The common ingredients for commitment are

(1) devotion to a rigorous daily exercise regimen,
(2) feeling separate from or superior to nonathletes,
(3) development of pre- or post-workout rituals (e.g., meditation),
(4) associating primarily with other athletes, and
(5) food preferences and practices that make having meals with others (especially nonathletes) problematic.

Step 5—constructing the data spreadsheet—is reported in table 7-1. Note that not all five conditions are shared by all twenty cases. In fact, the only condition shared by all twenty athletes is a devotion to a rigorous daily exercise regimen (exercise). However, the other four conditions are widely shared: 13/20 have a feeling of separateness (feel); 14/20 practice workout rituals (rituals); 13/20 associate primarily with other athletes (assoc); and 16/20 have distinctive food preferences or habits (food). Step 6, coding the outcome, is implemented in the last column of table 7-1 and affirms that all twenty athletes have maintained commitment for a substantial period of time (commit).

The next step (step 7) is to convert the data matrix into a truth table, which shows the different combinations of conditions found in the data spreadsheet, along with the number of cases displaying each combination. With five conditions, there are thirty-two logically possible combinations of conditions; only eight combinations have empirical instances, ranging in frequency from one to four athletes. Looking across the rows of the truth table, it is clear why diversity is limited—all rows have at least three of the five ingredients present.

Tables 7-2 and 7-3 display the truth table before and after the implementation of the interpretive coding of antecedent conditions. Recall from chapter 6 that interpretive inferences are central to the application of AI. Rather than using “presence versus absence” dichotomies, AI can utilize a different binary opposition: “contributing versus irrelevant.” The researcher uses her substantive and theoretical knowledge to determine which side of each presence/absence dichotomy is a contributing condition and defines the other side as irrelevant to the outcome in question. Assume, in this example, that the researcher interprets each of the five conditions in the truth table as contributing to the outcome when present (equaling 1), and as irrelevant otherwise. Accordingly, the zeros in each of the five condition columns are recoded to dashes (signifying irrelevance). The “raw” truth table is shown in table 7-2; the recoded truth table is shown in table 7-3.[4]

TABLE 7-1 Hypothetical data on committed Olympic athletes

| | Devotion to exercise (exercise) | Feeling of separateness (feel) | Workout rituals (rituals) | Associates with athletes (assoc) | Separate food (food) | Maintains commitment (commit) |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 0 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 0 | 0 | 1 |
| 4 | 1 | 1 | 1 | 0 | 1 | 1 |
| 5 | 1 | 1 | 1 | 0 | 0 | 1 |
| 6 | 1 | 1 | 0 | 1 | 1 | 1 |
| 7 | 1 | 0 | 0 | 1 | 1 | 1 |
| 8 | 1 | 0 | 1 | 1 | 0 | 1 |
| 9 | 1 | 0 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 0 | 0 | 1 | 1 |
| 11 | 1 | 1 | 1 | 0 | 1 | 1 |
| 12 | 1 | 1 | 0 | 1 | 1 | 1 |
| 13 | 1 | 0 | 1 | 1 | 1 | 1 |
| 14 | 1 | 0 | 1 | 1 | 1 | 1 |
| 15 | 1 | 1 | 1 | 0 | 1 | 1 |
| 16 | 1 | 1 | 0 | 1 | 1 | 1 |
| 17 | 1 | 1 | 1 | 0 | 1 | 1 |
| 18 | 1 | 1 | 1 | 1 | 1 | 1 |
| 19 | 1 | 1 | 0 | 1 | 1 | 1 |
| 20 | 1 | 0 | 1 | 1 | 0 | 1 |

TABLE 7-2 ”Raw” truth table based on data in table 7-1

| Exercise | Feel | Rituals | Assoc | Food | Number | Commit |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | 4 | 1 |
| 1 | 1 | 0 | 1 | 1 | 4 | 1 |
| 1 | 0 | 1 | 1 | 1 | 4 | 1 |
| 1 | 1 | 1 | 0 | 0 | 2 | 1 |
| 1 | 0 | 1 | 1 | 0 | 2 | 1 |
| 1 | 1 | 1 | 1 | 1 | 2 | 1 |
| 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 | 1 | 1 | 1 |

TABLE 7-3 Recoded truth table based on researcher's interpretive inferences

| Exercise | Feel | Rituals | Assoc | Food | Number | Commit |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | – | 1 | 4 | 1 |
| 1 | 1 | – | 1 | 1 | 4 | 1 |
| 1 | – | 1 | 1 | 1 | 4 | 1 |
| 1 | 1 | 1 | – | – | 2 | 1 |
| 1 | – | 1 | 1 | – | 2 | 1 |
| 1 | 1 | 1 | 1 | 1 | 2 | 1 |
| 1 | 1 | – | – | 1 | 1 | 1 |
| 1 | – | – | 1 | 1 | 1 | 1 |

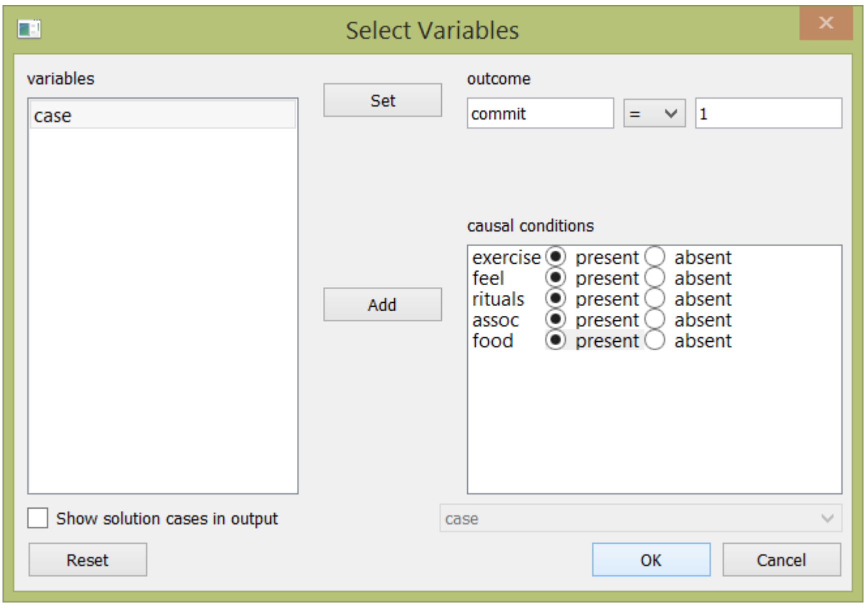


FIGURE 7-1. Dialogue box generating table 7-3.

While it is possible to recode fsQCA's truth table spreadsheet manually, the dialogue box for generalized AI enables the user to specify interpretive inferences, which in turn enables automated recoding of the truth table spreadsheet. Figure 7-1 shows the dialogue box that generated table 7-3.

The truth table is now ready for logical minimization (step 9).[5] After clicking "Run," minimization of the truth table yields the following recipes for commitment:

exercise•feel•rituals + exercise•rituals•assoc +
exercise•feel•food + exercise•assoc•food → commit

Note that the arrow indicates the superset/subset relation, a multiplication sign indicates the logical term *and* (combined conditions), a plus sign indicates the logical term *or* (alternate combinations of conditions), and a tilde indicates *not* (set negation). Altogether, there are four recipes for sustained commitment and only one common ingredient across the four: devotion to a daily exercise regimen (exercise).

At first glance, these results do not seem consistent with one of the core goals of AI, which is to identify shared antecedent conditions. However, it is important to recall the strategies outlined in chapter 2 for reconciling disconfirming evidence. One important strategy is to increase the scope of antecedent conditions, so that disconfirming cases are embraced (see esp. table 2-2). Using chapter 2's terminology, the goal is to move the disconfirming cases from cell *a* (outcome present, cause absent) to cell *b* (outcome present, cause present) by using logical *or* to join two or more closely related conditions (step 10).

Consider, for example, the condition "associates primarily with other athletes" (assoc). Referring back to table 7-1, seven athletes do not display this condition. However, these seven athletes all display a strongly related condition, "feeling separate from or superior to nonathletes" (feel). In fact, all twenty athletes display one or both of these two related ingredients. If these two conditions can be considered alternate ways of satisfying a more general requirement, then they can be joined using logical *or*. The resulting "macro-condition" (Ragin 2000) can be interpreted as alternate ways of constructing a boundary between athletes and nonathletes, and it has an invariant connection with the outcome (commit). That is, the macro-condition ("boundary construction") is a shared antecedent condition for the outcome (commit). Both the macro-condition and the outcome are constant across the twenty cases.

Notice that this same connection exists between "workout rituals" (rituals) and "separate food" (food). Whenever one of these two conditions is absent, the other is present. And they are closely related to each other, in that both involve everyday practices that reinforce an identity as an athlete. Considering these two conditions separately, they both fail to satisfy classic AI's strict requirement of shared antecedent conditions. However, they can be joined using logical *or* to create a macro-condition that has an invariant connection with the outcome (commit).

The general picture that emerges from the assessment of closely linked conditions is that there are three shared antecedent conditions, not just one (devotion to an exercise regimen). The twenty committed athletes share

(1) devotion to a daily exercise regimen,
(2) construction of a boundary separating athletes from nonathletes, and
(3) everyday practices that reinforce identity as an athlete.

Two of the antecedent conditions are macro-conditions that can be satisfied in either of two ways. It is important to note that creating macro-conditions entails the conceptualization of conditions that are more abstract than their component conditions. For example, "everyday practices that reinforce identity as an athlete" is pitched at a higher level of abstraction than "workout rituals." In general, expressing findings at a higher level of abstraction enhances their portability to other empirical domains (Vaughan 1986). Summarized as an equation, the reformulated results are much more compact than the four-recipe truth table solution:

$$\text{exercise} \bullet (\text{feel} + \text{assoc}) \bullet (\text{food} + \text{rituals}) \rightarrow \text{commit}$$

Note that this alternate representation of the results also can be derived by factoring the four-recipe solution (an alternate implementation of step 10). More generally, the original four-recipe solution is presented in "sum-of-products" form; the logically equivalent, reformulated solution just derived is presented in "product-of-sums" form. Viewing the results of generalized AI in the latter format can provide important clues regarding the construction of macro-conditions. Appendix D shows how to convert a sum-of-products equation into its logically equivalent product-of-sums form using fsQCA.

## DISCUSSION

This chapter offers a detailed example of generalized AI applied to a set of cases that share the same outcome. Researchers, especially those involved in qualitative investigations, often confront the task of making sense of a set of instances of an outcome. Because the outcome does not vary, conventional quantitative methods are of little use. Likewise, without negative cases, QCA is of limited utility (as demonstrated in chapter 6). By contrast, generalized AI provides important tools for making sense of such cases. The most important tool, in this regard, is the use of knowledge-based interpretive inferences to convert conventional "presence/absence" binaries into "contributing versus irrelevant" binaries. This translation makes it possible to consider case profiles holistically, as combinations of contributing conditions.