3

Explaining Variation versus Explaining Outcomes

What explains variation in the level or probability of an outcome? And what explains the occurrence of an outcome—how it comes about? These are two very important questions for social scientists. While obviously connected and often conflated, they are also quite different questions, with different starting points for finding an answer. For the first question, the starting point is cases that are "at risk" of displaying an outcome. For example, the population of recent high school graduates is "at risk" of attending college. An analysis of a sample of such graduates would focus on the predictors of college enrollment. Thus, implicit in the first question is the task of specifying the population of "candidates" for a given outcome, along with the expectation that the candidates will vary in outcome (Ragin 1992). The starting point of the second question, by contrast, is cases that actually display the outcome (Goertz and Haggard 2022). The focus is on understanding a qualitative outcome—how something happens (e.g., the process of becoming a college student, conceived as a happening), not on assessing which cases display the outcome versus those that do not. Cases that do not display the outcome can provide relatively little useful information about how an outcome happens.

More generally, the first question (concerning variation in the level or probability of an outcome) is centered on the problem of prediction (e.g., predicting who will attend college), while the second question (explaining the occurrence of an outcome) is centered on the problem of understanding (e.g., understanding the process of becoming a college student). The two questions also differ with respect to the goal of interpretation in social research. Predicting an outcome requires causal or statistical inferences; explaining how something happens entails interpretive inferences.

The gulf separating these two basic types of questions is clearly apparent in macro-comparative research. Consider, for example, the study of social revolutions. To answer the "variation in the outcome" question, it is necessary to construct the set of plausible candidates for social revolution, a task addressed with considerable sophistication by Mahoney and Goertz (2004: 665–68). The goal is to ensure that there is indeed variation in the outcome (e.g., presence vs. absence of social revolution), as well as variation in the relevant predictors of revolution (state breakdown, peasant insurrections, etc.). In other words, the researcher must assemble a set of candidates for social revolution, embracing both "positive" (successful) and "negative" (unsuccessful) cases. By contrast, answering the "How does it happen?" question mandates in-depth analysis of actual occurrences of social revolution (e.g., Crane Brinton's classic 1938 study *The Anatomy of Revolution* and the bulk of Theda Skocpol's 1979 study *States and Social Revolutions*). The first step of the analysis is to locate good instances of social revolution; the second is to identify and evaluate their shared antecedent conditions. Thus, while the first question is addressed by matching variation in the outcome to variation in relevant causal conditions, answering "How does it happen?" begins by linking a constant (positive instances of the outcome) to other constants (their shared antecedent conditions).

Qualitative research, with its emphasis on in-depth knowledge of cases, is the natural home of researchers who ask "How does it happen?" Quantitative research is the natural home of researchers who ask "What explains variation in the outcome?" Again, both questions are important, but they differ fundamentally. While answers to the first question have implications for answers to the second, and vice versa, it is unreasonable to expect consistency or even complementarity between the two types of analysis. After all, they address different questions. A simple example: Skocpol (1979) argues that state breakdown is a shared antecedent condition for social revolution—it is a constant across the cases she studied and clearly was a shared antecedent condition. However, state breakdown is experienced by many negative cases of social revolution as well. Thus, as an "independent" variable, it is a relatively poor predictor of social revolution, due to its weak correlation.

Neither approach to empirical evidence is inherently flawed or incorrect. The two approaches are simply different in both their starting points and their protocols for establishing and interpreting causal connections. However, substantial tension, if not outright rancor, separates practitioners of the two approaches. From the viewpoint of the quantitative approach, researchers who look only at positive cases are guilty of "selecting on the dependent variable." By contrast, from the viewpoint of the qualitative approach, and especially that of analytic induction (AI), quantitative researchers too often rely on given, taken-for-granted populations and may inadvertently pad their analyses with theory-confirming, but irrelevant, negative cases. I will address these two issues in turn.

SELECTING ON THE DEPENDENT VARIABLE

AI starts out with an interest in specific phenomena, qualitative outcomes, or happenings. At first, the conceptualization of the phenomenon to be explained is fluid and open to revision and reformulation. The usual expectation is that the phenomenon will become more completely specified as more is learned, usually through in-depth research at the case level. Thus, the initial focus is often on "good" instances of the qualitative outcome in question, and there is a back-and-forth between the identification of "good" instances and the specification of the nature of the phenomenon (Goertz 2017). At a formal level, the research focus is often on a specific category of phenomena, its constituent features, and relevant antecedent conditions and processes. After establishing "what it is," researchers focus on "how it happens." Similarities across instances of the phenomenon in question are a key focus in research of this type.

From the viewpoint of conventional quantitative research, the approach just sketched may seem ludicrous. First of all, the explanandum is more or less the same across all instances. Thus, the "dependent variable" does not vary, at least not substantially, and, accordingly, there is little or no "variation" to explain. Second, because the qualitative researcher has selected cases that have a limited range of values on the outcome (i.e., the researcher has selected on the dependent variable), correlations between antecedent conditions and the outcome are necessarily attenuated, which leads, in turn, to Type II errors (i.e., accepting the null hypothesis and concluding erroneously that hypothesized antecedent conditions are irrelevant to the outcome).

In *Designing Social Inquiry: Scientific Inference in Qualitative Research*, King, Keohane, and Verba (1994: 126–32) strongly discourage selection on the dependent variable. Their demonstration of the issue can be seen in figure 3-1, which reports hypothetical raw data showing the relation between the number of accounting courses taken by MBA students and their annual incomes after completing the degree. Two regression lines are plotted: a solid line showing the relationship for the whole sample, and a dashed line showing the relationship for graduates with incomes over $100,000. The authors' point is that the dashed line demonstrates the problem of selecting on the dependent variable—which, in this example, involves restricting the analysis to MBA graduates earning more than $100,000 annually. The dashed line is much flatter than the solid one, indicating lower income returns per number of accounting courses than in the full sample. They conclude that selecting on the dependent variable attenuates relationships and that the researcher who selects on the dependent variable may overlook important connections.

Viewed from the vantage point of AI, however, the "problem" of selecting on the dependent variable evaporates. Selecting on the high earners and then exploring their shared antecedent conditions, especially their academic backgrounds, would quickly lead to the conclusion that almost all high earners completed three or four accounting courses as MBA students. In fact, 82 percent of the high earner points in the figure

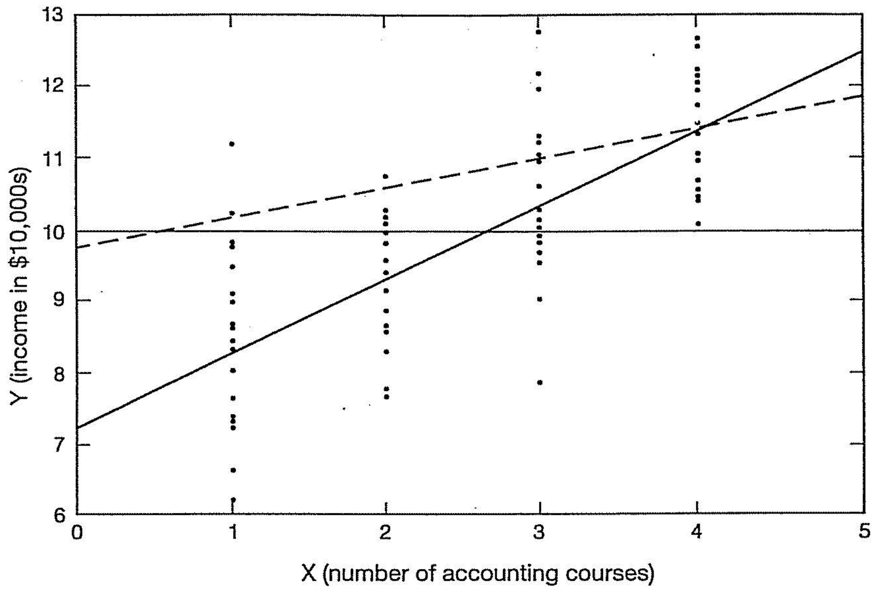

FIGURE 3-1. Recent MBA income levels plotted against number of accounting courses completed (from King et al. 1994: 131).

reside in the upper-right portion of the plot (three or four accounting courses completed). Thus, while selecting on the dependent variable may attenuate correlational relationships, it would not cause a qualitative researcher to miss this important connection. Only blind adherence to correlational methods would lead a researcher to overlook the strong connection between accounting courses and high income.

From the viewpoint of AI, the outcome or happening in this example is earning a high income (over $100,000). Completing three or four accounting courses as an MBA student is a widely shared antecedent condition for this outcome. By contrast, from the viewpoint of conventional quantitative research, the strong correlation between number of accounting courses and salary is clearly visible only when there are no restrictions on the range of the dependent variable.

IRRELEVANT NEGATIVE CASES

Answering the question "What explains variation in the outcome?" requires cross-case or longitudinal variation in the level, degree, or probability of an outcome. Thus—in contrast to answering "How does it happen?"—the set of cases with the outcome (or with sufficiently high levels of the outcome) cannot be used to circumscribe the entire set of cases relevant to an investigation. Instead, researchers must define the cases to be included in the analysis separately from the definition of the set of cases with the outcome. In other words, identifying the relevant population of cases and defining the dependent variable are separate tasks in conventional quantitative research. By contrast, these two tasks tend to be merged by researchers asking "How does it happen?"

For most quantitative research to proceed, cases must be drawn from a relevant and well-delineated population. The populations of conventional quantitative social science tend to be given or taken for granted. The key is that the population of relevant observations (i.e., cases) must be circumscribable. Often, however, the definition of the relevant population in quantitative research is contestable. Consider research on the causes of mass protest in Third World countries against austerity measures mandated by the International Monetary Fund (IMF) as conditions for debt restructuring. While it is a relatively simple matter to identify positive cases (i.e., countries with mass protest against IMF-mandated austerity), the set of relevant negative cases is somewhat arbitrary. Should the study include all less developed countries as candidates for IMF protest? Less developed countries with high levels of debt? Less developed, debtor countries with recent debt negotiations? Less developed, debtor countries subjected to IMF conditionality? Less developed, debtor countries subjected to *severe* IMF conditionality?

Each narrowing of the set of relevant cases, as just described, reduces the number of cases (*N*) available for quantitative analysis, which in turn undermines the possible utilization of advanced analytic and inferential techniques. Understandably, quantitative researchers generally avoid narrowly circumscribed populations. When *N* is small, standard errors tend to be large, and it is more difficult to generate findings that are statistically significant. For this reason, quantitative researchers often err on the side of being over-inclusive. In the example just presented, the preferred solution might be to include all less developed countries in the analysis and to use debt level and extent of IMF negotiations as "independent" variables.

While that solution seems plausible, at least on the surface, there is a world of difference between, on the one hand, using debt level and extent of IMF negotiations as independent variables, and, on the other, using these same variables to delimit the population of relevant candidates for mass protest against the IMF. These two uses are not only very different, from a statistical and mathematical point of view, but they call for very different analytic procedures. Using them as independent variables embraces all less developed countries as candidates for austerity protest; using them to delimit the relevant population shifts the focus to a relatively small but well-delineated subset of less developed countries—those that are clearly candidates for the outcome because of their high levels of debt and extensive IMF negotiations.

It is not generally recognized that boosting the sample size by casting a wide net carries with it an increased danger of Type I errors—erroneously rejecting the null hypothesis of no relationship. If *N* is artificially enlarged by including irrelevant negative cases (i.e., cases that are not plausible candidates for the outcome in question), then the correlations between causal and outcome variables are likely to be spuriously inflated (Mahoney and Goertz 2004). This artificial inflation occurs because irrelevant negative cases are very likely to have low scores on the independent variables and on the outcome variable, and thus will appear to be theory confirming, when in fact they are simply irrelevant. Correlational analysis is completely symmetrical in its calculation; therefore, a case with low (or null) values on both the causal and outcome variables is just as theory-confirming of a positive correlation as a case with high values on both. It is important to note, as well, that an artificially inflated *N* also increases the danger of Type I errors by reducing the size of estimated standard errors, which, in turn, makes statistical significance easier to achieve. For these reasons, it is important for quantitative researchers to ensure that all the cases included in an analysis are relevant—that they are plausible candidates for the outcome in question—especially in situations where the definition of candidate cases is contestable.

From the viewpoint of AI, the key focus is on instances of the outcome and on assessing their shared antecedent conditions. Once this task is complete, it is possible, though certainly not mandatory, to turn the analysis around and ask if there are cases that share the antecedent conditions—just identified—but not the outcome (i.e., cases in cell *d* of table 1-2). The guiding question regarding such cases is "What happened instead?" (e.g., what happened instead of IMF protest—martial law?), and very often there is a variety of alternate outcomes (see chapter 4). From an AI perspective, each alternate outcome is deserving of separate analytic attention.

In general, the greater the number of antecedent conditions shared by the positive cases, the smaller the number of cases that share the conditions but not the outcome. If there are no such cases, the researcher is left with only the original positive cases and their shared antecedent conditions. In effect, the researcher in this situation has established a pattern of results consistent with sufficiency because there are no cases that share the antecedent conditions but not the outcome. Also, as noted in chapter 1, classic AI's tendency to favor constitutive causal conditions, integral to the focal outcome, often guarantees that cell *d* (antecedent conditions present/outcome absent) will be void of cases.

ADDRESSING OUTCOMES THAT VARY BY LEVEL OR DEGREE

This chapter focuses on qualitative outcomes—"happenings" that are more or less binary (yes/no), such as attending college, protesting IMF-mandated austerity, and so on. The reader might infer that the arguments presented apply only to strictly qualitative outcomes, to the exclusion of the consideration of outcomes that vary by level or degree. However, the main arguments presented above regarding the study of happenings can be extended to include such outcomes. The application of fuzzy-set reasoning provides the way forward (see appendix B).

Consider, for example, the measurement of poverty and its calibration as a fuzzy set. The usual first step is to assess the composition of a household in terms of the number of adults and the number of children. This assessment provides the basis for specifying the poverty level for that household—the amount of income minimally necessary to support it. Next, the reported household income is divided by the poverty level for that household, to create each household's poverty ratio. A poverty score of 1 or lower indicates that the household is at or below the poverty level; a poverty ratio greater than 1 indicates that the household's income exceeds the poverty level for that household type. For example, a ratio of 1.5 would indicate that a household's income is 50 percent higher than the poverty level for that household.

While the evidence on household incomes and poverty levels is quantitative, the condition of being in poverty can be seen as a qualitative state once the ratio of income to poverty level is calibrated as a fuzzy set. With fuzzy sets, it is possible to assess the degree of membership of cases in sets, with membership scores ranging from 0 (fully out) to 1 (fully in). Three empirical anchors are used to calibrate the evidence so that it reflects qualitative concerns: the threshold for full membership in the target set, the crossover point (the point of maximum ambiguity in whether a case is more in or out of the set), and the threshold for full non-membership. For example, a poverty ratio of 3.0 (with household income three times the poverty level) could be used as the threshold value for being fully out of poverty. A ratio of 2.0 could be used to indicate maximum ambiguity in whether a household was more in or out of poverty, and a ratio of 1.0 could be used as a threshold value for full membership in the set of households in poverty. (See also chapter 9, especially figure 9-1, and appendix B.)

Essentially, the goal of fuzzy-set calibration is to create membership scores that reflect the substantive concerns of the researcher, which are implemented in the three values selected to shape the distribution of set membership scores. The next step in the analysis would be to select one or more qualitative breakpoints in the distribution of membership scores, consistent with the goals of the investigation. For example, the researcher might want to assess the antecedent conditions linked to full membership in the set of households in poverty and select cases that meet this threshold for further analysis. Do they share specific antecedent conditions? Alternatively, the researcher might choose a cutoff value of 0.75 membership, midway between full membership and the crossover point (i.e., 0.5—the point of maximum ambiguity regarding whether a case is more in or more out of the set in question). What antecedent conditions, if any, do these cases share? In short, the fuzzy-set metric offers multiple opportunities to operationalize specific qualitative concerns. Chapter 9 offers a detailed example of the implementation of multiple qualitative breakpoints using the fuzzy set metric.

DISCUSSION

The gulf between quantitative and qualitative social science is due, in part, to fundamental differences in the kinds of questions asked. This chapter has highlighted the methodological implications of two very different questions. Answers to "What explains variation in the level or probability of an outcome?" and "What explains the occurrence of an outcome?" have important implications for each other, but they require very different approaches to empirical evidence. The first question focuses equally on positive and negative cases and attempts to identify the best predictors, based on analyses of covariation with the outcome. The second question focuses on positive cases and attempts to identify their shared antecedent conditions.