ANALYTIC INDUCTION FOR SOCIAL RESEARCH

CHARLES C. RAGIN

Analytic Induction for Social Research

Analytic Induction for Social Research

Charles C. Ragin

UNIVERSITY OF CALIFORNIA PRESS

University of California Press
Oakland, California

Suggested citation: Ragin, C. C. *Analytic Induction for Social Research*. Oakland: University of California Press, 2023. DOI: https://doi.org/10.1525/luminos.159

Library of Congress Cataloging-in-Publication Data

Names: Ragin, Charles C., author.
Title: Analytic induction for social research / Charles C. Ragin.
Description: Oakland, California : University of California Press, [2023] | Includes bibliographical references and index.
Identifiers: LCCN 2022050606 (print) | LCCN 2022050607 (ebook) | ISBN 9780520393738 (paperback) | ISBN 9780520393745 (ebook)
Subjects: LCSH: Social sciences—Research. | Induction (Logic) | Analysis (Philosophy)
Classification: LCC H62 .R228 2023 (print) | LCC H62 (ebook) | DDC 300.72—dc23/eng/20230310

LC record available at https://lccn.loc.gov/2022050606
LC ebook record available at https://lccn.loc.gov/2022050607

ISBN 978-0-520-39373-8 (pbk. : alk. paper)
ISBN 978-0-520-39374-5 (ebook)

28 27 26 25 24 23
10 9 8 7 6 5 4 3 2 1

CONTENTS

ACKNOWLEDGMENTS

Many friends and colleagues at a variety of institutions directly enriched this book with their thoughtful comments and encouragement. I certainly will miss recognizing many who deserve thanks, and for that I apologize. That being said, my deep appreciation goes to David Byrne, Barry Cooper, Daniel Driscoll, Mary Driscoll, Gary Goertz, Jim Mahoney, Benoit Rihoux, and Claude Rubinson. I thank them all for their many sharp insights and helpful observations. Special thanks go to Peer Fiss for his detailed comments on the manuscript and, more generally, for his intellectual companionship of more than two decades. Special thanks also go to another long-term colleague, Edwin Amenta, who prompted me to fine tune my arguments while co-teaching graduate courses in comparative-historical methods.

The materials of this book have been presented, in various forms, in colloquia and workshops at the University of California, Irvine; University of Durham; University of Milan; University of North Carolina; Stanford University; Syracuse University; Case Western Reserve University; University of Oslo; University of Southern Denmark; Peking University; University of St. Gallen; University of Ljubljana; and University of Zürich/ETH. I am especially appreciative of Johannes Meuer and colleagues for organizing the annual QCA Expert Workshop in Zürich, a fabulous forum for exploring new ideas in set-analytic and configurational methods. I also thank Naomi Schneider and Summer Farah of UC Press for their help and enthusiastic support in getting this work to print. Sean Davey provided programming expertise and turned loosely connected algorithms and procedures into a coherent software package.

Over the years, students and colleagues have come to me with questions about their research designs. A common question concerns analytic techniques appropriate for situations where the cases included in a study all exhibit the same outcome. The textbook answer to this question is simple: "Don't do it." I decided, instead, to try to develop some useful tools, and this book is the culmination of my efforts. Thank you to all those who asked my advice. I learned a lot while trying to answer your questions.

I am deeply grateful to my family and especially to my lifelong partner, Mary, who has shared companionship with this work, supported me, kept encouraging me, and urged me over the finishing line. I dedicate this book to her.

Introduction

Social scientists ask diverse kinds of research questions. Usually, each such question calls for application of a specific analytic strategy to empirical evidence. For example, questions about the distribution of wealth in a population call for the analysis of variation in levels of wealth across a sample of households, using sociodemographic and other variables to predict levels. Analytic methods for the study of distributions are especially well developed in the social sciences today. Variation in a dependent variable (e.g., household wealth) is explained using variation in independent variables (e.g., race, ethnicity, immigration status, education). Social scientists have developed a vast array of variation-based analytic techniques, perfect for addressing questions about distributions.

But not all research questions are so lucky. Often, the research goal is to understand "how" a qualitative outcome happens by examining a set of cases that display the outcome. The distribution of that outcome in a sample drawn from a population will be relevant, but the empirical focus in determining the *how* of the outcome must rest on cases that display the outcome. Cases without the outcome—key evidence in the analysis of variation in the distribution of the outcome—can provide only very limited information regarding how the outcome happens. Restricting the analytic focus to cases that display the outcome, however, transforms the "dependent variable" into a *constant*—which precludes using the many variation-based analytic techniques that social scientists have developed. There is no readymade technique, comparable in sophistication to techniques that rely on a dependent *variable*, for the analysis of constants as outcomes.

Questions regarding how outcomes happen are quite common, though—especially in everyday discourse. Unfortunately, they are often recast by social scientists as questions about distributions. Imagine, for example, that instead of learning about the process of becoming a marijuana user by observing and interviewing users, Howard Becker (1953, 2015) had instead examined the distribution of marijuana use in a random sample drawn from a given population. Suppose he found high levels of use among musicians and certain other, related groups. While indirectly relevant to the *how* question, the finding does not address it head-on. To find out how one becomes a marijuana user, it is necessary to study users, focusing especially on their shared experiences in learning to use marijuana and on other widely shared antecedent conditions.

This book offers a straightforward methodology for the assessment of research questions regarding the antecedent conditions linked to qualitative outcomes. A typical qualitative study has a set of cases that display the outcome in question—the *focal* outcome—along with evidence on relevant antecedent conditions. The goal of the analysis is to identify antecedent conditions shared by cases with the focal outcome. Shared antecedent conditions, in turn, may be interpreted as "recipes" for an outcome, especially when they make sense as combinations of causally relevant conditions. In the end, the researcher explains a constant (the focal outcome) by way of other constants or near-constants (shared antecedent conditions).[1]

My approach to the analysis of systematic cross-case evidence on qualitative outcomes has deep roots in sociology in the form of a technique known as analytic induction (AI). AI was a popular research technique in the early decades of empirical sociology, beginning with the publication of Florian Znaniecki's (1934) *The Method of Sociology* (Tacq 2007). Exemplary AI studies include Alfred Lindesmith's (1947, 1968) *Addiction and Opiates*, Donald Cressey's (1953, 1973) *Other People's Money*, and Howard Becker's (1953, 2015) *Becoming a Marihuana User*. AI seeks to establish invariant (or "universal") conditions for qualitative outcomes, focusing exclusively on instances of the outcome and how it came about in each case.

As explained in chapter 1, early applications of AI used an especially strict version of the approach, which I call "classic AI." Classic AI (see also Becker 1998: 196–97) is strict in that it does not permit disconfirming cases, defined as cases where the outcome is present but one or more of the antecedent conditions specified in a working hypothesis is absent.[2] All instances of the outcome must be accounted for in some way, either by narrowing the definition of the outcome, thereby excluding disconfirming cases, or by respecifying the relevant antecedent conditions in a way that accommodates the disconfirming cases (see chapter 2). In fact, disconfirming cases are essential to classic AI because they provide raw material for refining the researcher's working hypothesis. They push the analysis forward. Very often, classic AI researchers seek out disconfirming cases, in order to refine their arguments, and in this way AI is akin to grounded theory's utilization of theoretical sampling based on inductively derived categories (Glaser and Strauss 1967; Katz 2001; Hammersley 2010).

However, as is so often the case with analytic methods, classic AI's strength is also its weakness. Accounting for every disconfirming case, as defined above, requires both in-depth knowledge of cases and substantial conceptual agility on the part of the researcher (see chapters 2 and 3). Besides, social phenomena are both heterogeneous and chaotic, data collection methods are imperfect, measures are crude and often contain known or hidden biases, revisits to research sites or subjects are often difficult or impossible, and coding mistakes are all too common (Katz 1983). One researcher's coding error is another researcher's disconfirming case, just as one ethnographer's observation of a wink is another ethnographer's observation of a blink. In principle, addressing disconfirming cases is a great way to fine tune a working hypothesis; in practice, however, it is often difficult to achieve satisfactory results (Becker 1958; Bloor 1978; Katz 1983).

Consequently, systematic applications of classic AI today are relatively rare. Instead, researchers interested in systematic cross-case evidence on qualitative outcomes routinely construct what I like to call *composite portraits* of their cases. For example, a researcher interested in the process of becoming a committed social movement activist might collect interview data on a diverse set of committed activists and attempt to identify common background characteristics and other shared antecedent conditions (see, e.g., Downton and Wehr 1998; Driscoll 2018). The researcher in this example would not expect to find every important background characteristic in every activist—as required by classic AI. Instead, the goal would be to identify background conditions that are widely shared by activists. The end product in this example would be an idealized composite portrait—an "ideal typic" (Weber 1949) activist who combines the major background characteristics identified by the researcher.

The composite portrait approach, as just described, has a lot in common with classic AI. The analytic scope is limited to cases that display the focal outcome. The research question asks, "How did the outcome happen, or come about?" The focus is on widely shared antecedent conditions, the expectation is that there are multiple antecedent conditions, and the researcher's goal is to make sense of shared conditions as a formula or recipe for the focal outcome. In fact, the pivotal difference between classic AI and the composite portrait approach just described is classic AI's insistence on identifying *invariant* antecedent conditions. For these reasons, it is appropriate to refer to the composite portrait approach as "generalized AI." It is generalized in the sense that it is a flexible adaptation of AI to the chaotic and capricious nature of social phenomena and to the many practical challenges of establishing invariant relationships.

As a substitute for classic AI's invariance requirement, generalized AI attends to frequency criteria. That is, the researcher attempts to identify *widely shared* antecedent conditions, not universally shared conditions. Thus, "enumerative" criteria—simple counts and proportions, for example—are utilized, but they are used to gauge the consistency of antecedent conditions, not to assess bivariate or multivariate relationships (Goertz and Haggard 2022). The latter would require an outcome that varies across cases, which AI eschews. Evaluating the generality of antecedent conditions across a range of positive cases—generalized AI's core procedure—is essentially an assessment of the "consistency" of set-theoretic relations

TABLE I-1 Contrasts between generalized analytic induction and conventional variable-oriented research

	Generalized analytic induction	Conventional variable-oriented research
Outcome	Constant across cases	Varies across cases
Focus	Causal formula or “recipe” based on shared antecedent conditions	Net effects of independent variables on a dependent variable
Scope of analysis	Cases with the outcome	A given population or defined set of candidates for the outcome
Negative cases	Not directly relevant	Essential
Explanatory template	Constants explain constants	Variables explain variables
Case selection	Diverse set of instances of the outcome	Representative sample drawn from a population or defined set
Research question	How the outcome happens	Relative effects of independent variables on the distribution of an outcome

(Ragin 2008: chaps. 1–3). Thus, generalized AI is best understood as a set-analytic technique, not a correlational one.

As an approach to social research, generalized AI differs fundamentally from conventional, variation-based approaches. The important contrasts between the two approaches are summarized in table I-1. As noted previously, generalized AI’s outcome is a constant—the set of cases displaying the outcome in question. While most such outcomes are qualitative in nature, it is possible as well to base the analysis on cases that meet a specified threshold of a quantitative variable (e.g., an income level signaling that an individual is well-off—see chapter 9). Conventional variable-oriented research, by contrast, is centered on the task of explaining variation in a dependent variable, focusing on the net effects of independent variables (Ragin 2006b). Another key contrast is the role of “negative” cases—that is, cases that fail to exhibit the focal outcome. Such cases are not considered disconfirming according to generalized AI. Instead they are considered instances of an alternate outcome and therefore are the focus of a separate analysis altogether. By contrast, negative cases in conventional quantitative research are valued for their contribution to variation in the dependent variable.

It is important to point out that unlike much variable-oriented research, generalized AI is not inferential. Instead, it is primarily descriptive and is best understood as an aid to causal interpretation. It can be used in conjunction with other analytic methods, including conventional quantitative methods, by providing results in the form of causal recipes. Conventional quantitative methods focus primarily on isolating the separate, net effects of “independent” variables, not on their conjunctural impact. This aspect undermines the utility of conventional quantitative methods for causal interpretation, which often involves a focus on recipe-like combinations of conditions.

The application of generalized AI's core procedure is ubiquitous in social research, especially in qualitative work (Bernard et al. 2017). It's obvious that a lot can be learned from exploring the antecedent conditions shared by positive instances of an outcome (Goertz and Haggard 2022). Unfortunately, most applications of the core procedure are unsystematic and ad hoc. Only rarely do researchers quantify their assessments, and seldom do they explore *combinations* of conditions linked to an outcome. My main argument in this book is that there is a lot to be gained from systematizing generalized AI as a set-analytic method. In the chapters that follow, I make the case for treating generalized AI as a formal technique (see also Ragin and Amoroso 2019: 112–17).

OVERVIEW

Part I of this book (chapters 1–4) examines classic AI and addresses basic research-design issues associated with its use. Chapter 1 introduces the method, detailing its logic, describing it as a series of steps, and reviewing some exemplary applications. I also touch on the controversy stirred by classic AI, especially following W. S. Robinson's (1951) critique in the *American Sociological Review*. Along the way, I compare correlational approaches to causation with set-analytic approaches and describe AI's contrasting approach to two very different kinds of "disconfirming" cases: those that display the antecedent conditions specified in a working hypothesis but not the outcome, and those that display the outcome but not the hypothesized antecedent conditions.

Chapter 2 offers a thorough discussion of AI-based methods for addressing disconfirming cases—that is, instances of an outcome that fail to display the antecedent conditions specified in the researcher's working hypothesis. There are two main strategies for reconciling such cases. One is to narrow the definition of the outcome so that disconfirming cases are excluded. The other is to expand the breadth of the working hypothesis in a way that accommodates the disconfirming cases. It is also possible to address disconfirming cases by developing outcome subtypes or through the specification of appropriate scope conditions.

Chapter 3 examines the methodological implications of two very different types of research questions. On the one hand, what explains variation in the level or probability of an outcome? On the other, what explains the focal outcome's occurrence—how it comes about? The key is that the first question is focused on the distribution of an outcome in a given sample or population, while the second is focused more or less exclusively on positive instances of the outcome. These two different ways of conducting social science have spawned widespread disagreement and controversy. In one camp, researchers who seek to explain variation reject the other side's "selection on the dependent variable." Meanwhile, in the opposing camp, researchers focused on understanding how instances of an outcome happen reject a common practice of the other side: boosting the sample size of cases by casting a wide net, thereby running the risk of including irrelevant cases.

Chapter 4 contrasts three approaches to the analysis of dichotomous outcomes: conventional quantitative analysis, qualitative comparative analysis (QCA), and AI.[3] The three approaches 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. 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 consistently display a given outcome. By contrast, negative cases of the outcome play no direct role in AI, which separates the analysis of positive cases from the analysis of negative cases. In this "fully asymmetric" approach, negative cases are viewed as positive cases of one or more alternate outcomes.

Part II (chapters 5–10) offers a detailed presentation of generalized AI. Chapter 5 introduces Part II by briefly summarizing key differences between generalized AI and classic AI. Chapter 6 describes an essential feature of generalized AI: its reliance on "interpretive inferences" based on substantive and theoretical knowledge. Interpretive inferences transform presence-versus-absence conditions into contributing-versus-irrelevant conditions. For example, substantive knowledge indicates that being educated contributes to avoidance of poverty. On the basis of this knowledge, a researcher would bypass consideration of "not being educated" as a condition for avoiding poverty. If a person who has successfully avoided poverty is uneducated, then their lack of education is eliminated as a possible contributing condition of their avoidance of poverty. This feature of AI contrasts sharply with QCA's configurational logic, which requires both sides of every presence/absence condition to be treated equally.[4] Configurational logic dictates that the researcher entertain the possibility that not being educated could contribute to successfully avoiding poverty.

Using hypothetical data on Olympic-caliber athletes, chapter 7 offers a step-by-step application of generalized AI to the analysis of a set of cases that share the outcome "sustained commitment." Many researchers, especially those who conduct qualitative investigations, are routinely tasked with making sense of a set of instances of an outcome. Because the outcome in question does not vary, a conventional quantitative approach is of little use here—as is, without negative cases, QCA (as demonstrated in chapter 6). By contrast, generalized AI provides important tools for making sense of such cases.

A reanalysis of data published in Jocelyn Viterna's (2006) study of women's mobilization into the Salvadoran guerrilla army is the focus of chapter 8. Viterna applies key principles of generalized AI in her pathbreaking study. She distinguishes five different outcomes—three distinct paths to guerrilla activism (politicized, reluctant, and recruited) and two non-guerrilla paths (collaborators and nonparticipants). Rather than define the analysis as a binary contrast between the three guerrilla paths versus the two non-guerrilla paths, she focuses instead on the separate conditions linked to each of the five outcomes. She views each of the outcomes as worthy of separate analytic attention and thereby avoids conventional dichotomization of the outcome as "guerrilla versus non-guerrilla." This feature of her study, along with several others, aligns well with generalized AI.

Chapter 9 tackles the problem of bridging generalized AI and conventional quantitative analysis. It demonstrates that generalized AI can be usefully applied to conventional quantitative data. Because generalized AI is fundamentally descriptive in nature, it can complement findings derived using conventional quantitative methods. The demonstration of generalized AI uses data on Black females from the National Longitudinal Survey of Youth (NLSY), 1979 sample. The focus is on two outcomes, analyzed separately: membership in the set of respondents in poverty, and membership in the set of respondents well out of poverty. The results are asymmetric, with different conditions linked to the two outcomes.

The final chapter summarizes the essential features of generalized AI, as presented in this book. The listed features range from generalized AI's orientation as a research approach to practical procedures involved in applying the method.

A NOTE ON THE CONCEPT OF CAUSATION

The primary objective of this book is to provide tools that aid causal *interpretation*. Tools for causal *inference*, by contrast, are beyond its scope. More generally, the approach to causation advocated in this work is based on the regularity theory of causation. According to this theory, causation is indicated by an invariant connection between cause and outcome, which is also a concern of classic AI as described in this book. Classic AI adheres to John Stuart Mill's version of regularity theory, in particular his method of agreement, which selects on instances of an outcome and seeks to identify their shared antecedent conditions.

The relation between antecedent conditions and outcomes is set-theoretic in nature: instances of the outcome constitute a subset of instances of the antecedent conditions. This subset relation is evident, for example, whenever instances of an outcome agree in sharing a causally relevant antecedent condition. Of course, perfect set relations are relatively rare in social research. Thus, this book emphasizes assessing the degree of consistency of empirical evidence with the subset relation in question and restricts the analytic focus to connections that are highly consistent.

The designation of conditions as *causally relevant* to an outcome is dependent on theory and knowledge, and thus open to contestation. The larger task of specifying the "true" causes of social phenomena is beyond the scope of this work, and indeed beyond the purview of most social science methodology. Usually, social scientists must be content with successfully identifying causally relevant antecedent conditions, which in turn are suggestive of causal mechanisms. The true test of any hypothesized antecedent condition is its relevance at the case level. It is at the case level that social researchers have the opportunity to observe and narrate causal processes and mechanisms (Goertz and Haggard 2022). Thus, establishing regularities is essential, but it is not the whole story. Whenever possible, researchers should complement the identification of regularities with confirmatory process tracing at the case level.

PART ONE

The Logic of Analytic Induction

1

Classic Analytic Induction

Analytic induction (AI) was a popular technique in U.S. sociology during the early decades of empirical social research. The method was first formalized by Florian Znaniecki (1934) in his book *The Method of Sociology*. Znaniecki believed AI to be more scientific than "enumerative induction" (known today as correlational analysis) because of AI's emphasis on "universals"—invariant connections between antecedent conditions and outcomes (Tacq 2007). The basic idea was that the researcher should pinpoint antecedent conditions uniformly shared by instances of an outcome. Thus, the method focuses on *positive* instances of an outcome and attempts to provide an account of the outcome's etiology based on an analysis of shared antecedent conditions.[1]

AI's focus on the antecedent conditions shared by instances of an outcome is rooted in John Stuart Mill's method of agreement. He argues that if two or more instances of the phenomenon under investigation have only one circumstance in common, that one circumstance is the cause (or effect) of the given phenomenon (Mill 1967). In short, his method of agreement dictates close inspection of the antecedent conditions shared by instances of the phenomenon under investigation. While he frames the definition of the method of agreement in terms of a single shared condition ("only one circumstance"), his argument can be easily extrapolated to situations where there is more than one shared circumstance. Together, multiple shared conditions can be understood as contributing causes in situations where their combination is seen as a causal formula or recipe.

Both AI and Mill's method of agreement are formalizations of a very common technique for deriving empirical generalizations. Humans look for connections in everyday experiences and draw conclusions from repeated observations. For example, the observation that I must leave home for work by 7:00 a.m. in order to avoid heavy automobile traffic is an empirical generalization, based on repeated experiences. A consistent antecedent condition for the avoidance of heavy morning traffic is on-time departure for my commute to work. Of course, the consistency of the connection may be far from perfect, but still consistent enough to guide my behavior.

While commonplace, the search for antecedent conditions shared by positive instances can be the basis for prizewinning research. Consider, for example, Elinor Ostrom's (1990) research reported in *Governing the Commons*. Her main target was a widely held view of common-pool resources: that, absent state oversight and management, such resources are likely to be abused and rendered unsustainable through overuse.[2] To counter this view, Ostrom studied a variety of *self-governing* common-pool resources, where there were successful collective efforts to achieve sustainability, orchestrated by the surrounding communities. Ostrom observed that these *positive* cases shared a number of characteristics, including, for example, rules that clearly defined who gets what, good conflict-resolution methods, users who monitor and punish violators, and so on. In short, she established important preconditions for community-based resource sustainability based on her analysis of positive cases. She won the Nobel Prize in Economics for her research.[3]

Another example of this strategy in comparative research is Daniel Chirot's *Modern Tyrants* (1996). Examining thirteen tyrants, drawn from diverse settings, Chirot writes that "tyrannies have come to power in states both big and small; in rich industrial and very poor agrarian societies; in countries with many centuries of statecraft in their tradition, and in brand new ones; in culturally united nations with a firm sense of identity, and in ethnically split states with almost no basis for common solidarity" (Chirot 1996: 403). He asks, "What generalizations can be drawn from these thirteen sad and diverse histories?" While acknowledging that his conclusions are probabilistic in nature (418), he offers eight generalizations based on his study of thirteen tyrants, noting, for example that "the more chaotic the economy and political system, the more they seem to be failing, the more likely it is that a tyrant will emerge as a self-proclaimed savior" (409).

AI is often overlooked as a formal technique because it is simultaneously ubiquitous and rare. It is ubiquitous because it is based, as just described, on a very common method of generalizing about empirical regularities from equivalent observations (Bernard et al. 2017). Why formalize or even cite a method that seems like common sense? By contrast, applications of *classic* AI are somewhat rare because of its requirement that researchers demonstrate *invariant* connections between outcomes and antecedent conditions. All exceptions to working hypotheses must be addressed and resolved. As detailed in this and subsequent chapters, this feature of classic AI mandates both in-depth knowledge of cases and conceptual agility on the part of researchers. For some analysts, the invariance requirement dictates a determined pursuit of disconfirming cases—positive cases of the outcome that do not exhibit the antecedent conditions specified in a working hypothesis (Katz 1983; Denzin 2006; Athens 2006).

This book, while building upon classic AI, ultimately relaxes several of its defining features in order to lay the foundation for *generalized* AI. For example, the invariance requirement is unrealistic for the work of many researchers and research projects. Typically, researchers have a fixed set of collected data and little or no opportunity to return to the cases for more evidence or to seek out new cases that might challenge a working hypothesis. Another example: classic AI has little use for frequency criteria because a single disconfirming case can torpedo a working hypothesis. For most empirically minded social scientists, however, the weight of the empirical evidence matters, and frequency criteria are considered not only informative, but often decisive (Goertz and Haggard 2022; Miller 1982).

This chapter provides an extensive discussion of classic AI, focusing on the logic of the approach. First, I examine several classic examples of the approach and then formulate the method as a series of steps. Classic AI is both dynamic and iterative. It is a research approach that builds empirical generalizations on the basis of in-depth case knowledge. Second, I examine classic AI's understanding of causation, contrasting it with more conventional forms of analysis.

SOME EXAMPLES OF CLASSIC AI

Early, exemplary studies utilizing classic AI include Alfred R. Lindesmith's *Addiction and Opiates* (1947 [titled *Opiate Addiction*], 1968), Donald R. Cressey's *Other People's Money* (1953, 1973), and Howard S. Becker's *Becoming a Marihuana User* (1953, 2015). All three studies offer detailed portrayals of AI as a research process that builds a coherent argument based on in-depth analysis of cases.

Drawing on his interviews with more than sixty addicts, Lindesmith attempted to identify the antecedent conditions linked to opiate addiction. He argued that users become addicts only when they consciously use the substances to diminish the effects of withdrawal (Lindesmith 1968: 191). In other words, there is an important cognitive component to opiate addiction. Addicts must realize that this is why the effects are happening, and that no other physical ailment explains the painful withdrawal symptoms (1968: 191). If they do not attribute the withdrawal as such to their opiate use, and they believe that some other physical deficiency causes the side effects, they do not become addicts. Lee and Fielding (2004) summarize Lindesmith's argument as a specification of the process of becoming an addict: these individuals (a) use an opiate; (b) experience distress due to withdrawal of the drug; (c) identify or recognize the symptoms of withdrawal distress; (d) recognize that these symptoms will be alleviated if they use the drug; and (e) take the drug and experience relief (see also Becker 1998: 197).

The purpose of Cressey's study was to look at the sequence of conditions that lead an individual in a trusted financial position to embezzle money (Cressey 1973: 12). He gathered interview data from 210 convicted embezzlers, asking them about their experiences before, during, and after they were caught violating trust. After a lengthy process of reformulating hypotheses, identifying themes, and connecting them to a general concept, Cressey concludes that there are three necessary conditions: the individual (1) perceives that a personal, non-shareable financial problem has occurred, (2) rationalizes a reason for taking entrusted funds, and (3) believes that this is the only way to solve the non-shareable problem (1973: 139). It is important to point out that Cressey allowed for the possibility that a necessary condition could be satisfied in more than one way. For example, he identified three circumstances in which embezzlers "rationalized" their behavior: (1) they needed or wanted to borrow money, (2) they felt that the funds belonged to them, or (3) they felt it was a one-off situation (1973: 101–12).

Becker interviewed fifty recreational marijuana users in an effort to specify the process of becoming a user. His stated goal was to document the necessary conditions for recreational marijuana use (Hammersley 2011). He argued that there are three universal conditions or steps that must occur, at some point, in order for one to become a recreational marijuana user: (1) smoking it properly to induce a high, (2) recognizing and understanding the effects caused by the drug, and (3) learning "to enjoy the sensations" (Becker 1953: 242). Without satisfying all three conditions, an individual will not be able to become a recreational marijuana user. Becker draws an important distinction between those who use marijuana for pleasure and those who use the drug, but not for pleasure, and restricts his account of marijuana use to the former.

Based on these early applications, it is clear that AI is a discovery-oriented, abductive tool (Diesing 1971; Tavory and Timmermans 2014). It is also clear that because of its requirement of causal invariance, applications of classic AI tend to focus on antecedent conditions that are proximate to the outcome in question. Indeed, the antecedent conditions identified in these exemplary AI studies could be seen as constitutive of their outcomes (Turner 1953), which in turn suggests, to Lindesmith (1952), that the conditions are not only necessary but also sufficient.

In *Poor People's Lawyers in Transition*, Jack Katz (1982) offers a detailed illustration of AI's dynamic nature, especially the process of "double fitting" the conceptualization of causally relevant conditions with the conceptualization of the outcome. More recent applications of classic AI include the work of Hicks (1994), Gilgun (1995), Monaghan (2002), and Bansal and Roth (2000). In political science, there are several notable examples of work utilizing principles of AI. In addition to Chirot's *Modern Tyrants*, these include Guillermo O'Donnell and Philippe Schmitter's *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies* (1986), Crane Brinton's *The Anatomy of Revolution* (1938), O'Donnell's (1973) work on the origins of bureaucratic-authoritarian regimes in South America, and Juan Linz and Alfred Stepan's *The Breakdown of Democratic Regimes* (1978).

CLASSIC AI: STEPS

Various authors (e.g., Robinson 1951; Cressey 1973; Hammersley and Cooper 2012: 131–32) have attempted to capture classic AI's dynamic, iterative character by formalizing the method in terms of a series of steps:

1. Specify the outcome to be explained. Typically, the outcome is qualitative in nature. For example, it might be a "happening" or an occurrence like becoming an embezzler (Cressey 1973) or becoming a marijuana user (Becker 1953). The happening also can be meso- or macro-level (Katz 2001)—for example, episodes of mass protest against an authoritarian regime.
2. Collect evidence on a number of cases in which the outcome occurred. Usually, these are very clear instances of the outcome in question (Goertz 2017: 63–66). Some versions of classic AI (e.g., Lindesmith 1968) restrict the initial investigation to a single case, then add more cases one at a time (see also Robinson 1951; Lee and Fielding 2004). This restriction ensures that each case will be subjected to an in-depth assessment. However, for many investigations, this restriction is neither feasible nor warranted.
3. Identify the causally relevant antecedent conditions shared by these initial instances of the outcome. Formulate a working hypothesis on the basis of observed commonalities. Existing theory and substantive knowledge regarding relevant causal conditions for the outcome serve as preliminary guides. The commonalities identified by the researcher must make sense, on either substantive or theoretical grounds, as antecedent conditions.
4. Seek out and collect evidence on additional instances of the outcome. It is more important that the selected cases are diverse than that they are representative of a population (Goertz and Mahoney 2012: 182–85). Researchers should identify and study instances of the outcome that challenge their working hypothesis.
5. If cases are found that challenge a working hypothesis, then either the antecedent conditions or the outcome (or both) must be reformulated in some way. Typically, if the outcome is reformulated, its scope is narrowed so that the nonconforming cases are excluded from the purview of the working hypothesis. If the antecedent conditions are reformulated, the causal argument is altered so that the nonconforming cases are embraced in some way, typically through a strategy of conceptual realignment. In either situation, the process of reformulation should be both public and transparent.
6. Continue steps 4 and 5 until the evidence derived from additional instances no longer prompts reformulations of the working hypothesis or its empirical scope. The research has reached a point of theoretical saturation and an invariant connection has been established.

As this summary of classic AI's steps makes clear, AI's "dependent variable" is not a variable, but a constant; it is an outcome that is more or less the same across selected cases. This type of analysis is beyond the purview of conventional quantitative methods, which are focused on explaining variation in dependent variables by using variation in independent variables. Worse yet, examining only positive cases is viewed in the quantitative literature as an extreme form of "selecting on the dependent variable"—a great sin to be avoided, according to some authors (e.g., King et al. 1994). AI has little use for the analysis of the covariation of variables. Instead, the goal is to explain a constant, the outcome, with other constants—their shared antecedent conditions. The end result is a specific type of empirical generalization, one that is set-analytic, as opposed to correlational, in nature.

For example, the observation that social revolutions share peasant insurrections as an antecedent condition (Skocpol 1979) casts social revolution as a subset of instances of peasant insurrection. In this example, a *connection* between two sets (the set of countries with social revolutions and the set of countries with peasant insurrections) provides the basis for an empirical generalization. This observed connection stands on its own, without reference to variation in the presence versus the absence of either social revolution or peasant insurrection. Instead, the presence of peasant insurrection is linked to the presence of social revolution. It does not matter that there are many instances of peasant insurrection not linked to social revolution. By contrast, most empirical generalizations in the social sciences today are based on correlations between variables. For example, a researcher might offer an empirical generalization based on a positive correlation between social inequality and social unrest. In general, social scientists have not acknowledged connections between sets as a separate type of empirical generalization, distinct from those based on covariation.

THE LOGIC OF AI

AI was challenged as a technique for studying causation in 1951 by W. S. Robinson, in his article "The Logical Structure of Analytic Induction," published in the *American Sociological Review*, then and now the flagship journal of the discipline. His basic argument is that the method is fundamentally flawed because it can only identify necessary conditions, and therefore is not suitable for prediction. If it is used at all, it must be complemented with or followed by "enumerative induction" (i.e., correlational analysis) to certify that the causal factors identified using AI are in fact predictive (Miller 1982; Goldenberg 1993).

To fully grasp the substance of Robinson's critique, it is important to consider the essential differences between correlational analysis (Robinson's favored technique; see also Miller 1982) and set-theoretic analysis. The core principle of correlational analysis is the assessment of the degree to which two series of values parallel each other across comparable cases.[4] The simplest form is the 2 × 2 table

TABLE 1-1 Correlational approach to causation

	Cause absent	Cause present
Outcome present	Cell *a:* cases in this cell contribute to error	Cell *b:* many cases should be in this cell
Outcome absent	Cell *c:* many cases should be in this cell	Cell *d:* cases in this cell contribute to error

TABLE 1-2 Set-analytic approach to causation

	Cause absent	Cause present
Outcome present	Cell *a:* cases in this cell contradict necessity	Cell *b:* cases in this cell are consistent with both necessity and sufficiency
Outcome absent	Cell *c:* cases in this cell are not directly relevant to either necessity or sufficiency	Cell *d:* cases in this cell contradict sufficiency

cross-tabulating the presence/absence of a cause against the presence/absence of an outcome (table 1-1). Correlation is strong and in the expected direction when there are as many cases as possible in cells *b* and *c* (both count in favor of the causal argument, equally) and as few cases as possible in cells *a* and *d* (both count against the causal argument, again, equally).

Because cases in cell *c* are as hypothesis-confirming as cases in cell *b*, researchers must guard against including irrelevant cases in their analyses. Irrelevant cases would likely reside in cell *c* (cause absent/outcome absent) and thus spuriously confirm the researcher's hypothesis and contribute to a Type I error. In short, researchers who utilize the correlational template (which embraces the bulk of conventional quantitative social science) for their analyses must ensure that the cases they include are all valid candidates for the outcome in question (see chapter 3).

The set-analytic approach to this same 2 × 2 table differs substantially from the correlational approach (Miller 1982), as demonstrated in table 1-2. Each of the four cells has a different interpretation (Goertz 2017). The analytic focal point is cell *b*, which captures the cases that exhibit both the cause and the outcome (2017: 63–66). But is the cause a necessary condition for the outcome? If so, then cell *a* should be empty.[5] Is the cause sufficient for the outcome? If so, then cell *d* should be empty. Thus, the set-analytic approach to the 2 × 2 tabulation of outcome by cause is to separate the two causal relationships embedded in the table. After all, a cause can be sufficient but not necessary, and it can be necessary but not sufficient.[6] Note also that cases in cell *c* (cause absent/outcome absent) are not involved in either assessment. While cell *c* cases are integral to correlational analysis, computationally equal in importance to cases in cell *b* (cause present/outcome present), they play no direct role in the set-analytic approach.

Thus, two very different kinds of disconfirming cases are represented in table 1-2 (see also Ragin 2008). Cell *a* contains cases where the outcome is present, but the hypothesized cause is absent; cell *d* is the opposite—the hypothesized cause is present, but the outcome is absent. Robinson (1951) is correct in noting that classic AI focuses primarily on necessary conditions. Classic AI's central concern is the first row of table 1-2, especially the challenges to a working hypothesis posed by disconfirming cases in cell *a*.[7] Furthermore, addressing and reconciling cell *a* cases is the primary means of theoretical advancement in classic AI. Thus, the technique has little interest in cases that occupy cells *c* and *d*. Cases in cell *c* (cause absent/ outcome absent) are not directly relevant to the assessment of either necessity or sufficiency and thus can be safely set aside. Some AI researchers (e.g., Cressey 1973: 31) do utilize hypothetical cases in cell *c* by arguing that their cell *b* cases were in cell *c* before they experienced the relevant causal conditions associated with the outcome. In effect, these cases traveled from cell *c* to cell *b* once the right causal conditions were present.

The issue of disconfirming cases in cell *d* (cause present/outcome absent), however, deserves further attention. Cases in cell *d* could be seen as AI's blind spot, because it is standard AI practice to focus on cases with the focal outcome—meaning that cases in cells *c* and *d* are routinely bypassed. However, there are several factors to consider regarding AI's apparent disinterest in cell *d* cases:

1. It is important to note that AI focuses primarily on questions regarding how outcomes happen. As explained in detail in chapter 3, AI views outcomes as happenings and seeks to account for happenings in terms of their shared antecedent conditions. Cases in cell *d* fail to exhibit the focal outcome and thus can provide very little useful information regarding how it came about (see also point 6 below). Cases in cell *a*, by contrast, experienced the outcome but not the hypothesized causes and thus offer important raw material for clarifying the outcome's etiology.
2. From the viewpoint of AI, cell *d* cases experience a different outcome, compared to cell *b* cases. The cell *d* outcome is deserving of separate investigation, culminating in a specification of its etiology (Kidder 1981). In short, the cell *d* outcome, if there is one that is shared by these cases, should not be treated simply as instances of the absence of the focal outcome (i.e., as mere negative cases), but as instances of an alternate outcome that is worthy of separate analytic attention. For example, if Cressey (1973: 31) found that his cell *d* cases resorted to suicide, not embezzlement, once confronted with a non-shareable financial problem, that outcome would become the focal point of a separate investigation.
3. Typically, however, cell *d* cases display a wide variety of nonfocal outcomes and are thrown together only by the fact that they did not experience the focal outcome (see chapter 4). Of course, the researcher may choose to

document the different outcomes and identify the antecedent conditions specific to each, but this effort would be secondary to understanding the focal outcome, displayed by cell *b* cases.

4. Cases in the first row of table 1-2 (cells *a* and *b*) comprise a relatively well-defined and circumscribed set of cases—they are all instances of the focal outcome. The second row, which contains cases lacking the focal outcome, is not so well defined (see chapter 4). Presumably, cases in the second row are or were viewed as candidates for the focal outcome; otherwise, their inclusion in the analysis would not be justified. However, the definition of candidacy for the focal outcome may be arbitrary, which in turn makes the decision regarding which cases to include in the second row contestable (Ragin 1997, 2009). By focusing on the first row, AI bypasses the problem of circumscribing the set of valid negative cases—cases that might have experienced the outcome, but did not.
5. In general, AI focuses on shared antecedent conditions for an outcome. Very often, the "cause" in table 1-2 is not a single condition, but a combination or sequence of conditions. The greater the number of antecedent conditions the researcher is able to identify, the less likely there will be cases in cell *d*. In short, as more antecedent conditions are added to the mix, the number of cell *d* cases that meet them all may be correspondingly diminished. Full articulation of relevant antecedent conditions could easily lead to an empty cell *d*, which would provide evidence consistent with an argument of causal sufficiency (Ragin 2008: 17–23). The fact that cell *d* can be emptied of cases as the researcher specifies more antecedent conditions explains, in part, why Lindesmith (1952) responded to Robinson's (1951) critique by arguing that the conditions identified by classic AI were not just necessary, but necessary *and sufficient*.
6. Cases in cell *d*, if they exist, have a potentially useful role—they can help the researchers refine their articulation of the etiology of the focal outcome. Because cell *d* cases display the causal conditions but not the focal outcome, close inspection of cell *d* cases can lead to the identification of conditions that either neutralize one or more of the antecedent conditions manifested in cell *b* cases or block the outcome altogether. However, because cases in cell *d* are likely to be heterogeneous (see point 3 above) and their inclusion as negative cases contestable (see point 4), they may offer only limited analytic leverage.
7. Classic AI's invariance requirement tends to favor the identification of antecedent conditions that are proximate to and constitutive of the outcome. Turner (1953: 608) goes so far as to argue that applications of classic AI culminate in constitutive definitions of the outcome, not causal explanations. Consequently, any case that displays the antecedent conditions specified by

> the classic AI researcher may automatically display the outcome. As a result, cell *d* cases (condition present/outcome absent) may be extremely rare, if they exist at all.

Given these considerations, AI's apparent disinterest in cases in cell *d* is understandable. Note also that several of these considerations upend Robinson's (1951) critique of AI. His critique focuses on AI's inability to predict an outcome, based on its failure to take into account cases in the bottom row of table 1-2. But Robinson failed to consider (1) the goal of AI—to explain how an outcome happens, not to predict its distribution in a population or a sample; (2) the problematic nature of the definition of relevant negative cases—that it is often arbitrary and contestable; (3) the heterogeneity of negative cases—that they may include many alternate outcomes, each suggesting a possible avenue for further investigation; and (4) the fact that there may be no cases in cell *d*, due to a comprehensive specification of relevant antecedent conditions.

LOOKING AHEAD

Chapter 2 presents analytic strategies for addressing disconfirming cases, building on table 1-2 as a template. Altogether there are six main strategies, all focused on emptying cell *a* of cases. In general, the strategies are consequential from a set-analytic perspective because their goal is to document an invariant relationship between one or more antecedent conditions and an outcome. By contrast, these strategies typically yield only very modest gains from a conventional variable-oriented perspective.

2

Reconciling Disconfirming Cases

Classic AI rests on three main pillars: (1) focusing on positive instances of an outcome, (2) identifying their shared antecedent conditions, and (3) assessing the substantive and conceptual implications of disconfirming cases. This chapter describes key analytic strategies involved in implementing the third pillar.

Despite its name, AI is both inductive and deductive (Hammersley and Cooper 2012; Katz 2001; Manning 1982). For example, the identification of causally relevant antecedent conditions depends both on case knowledge (the inductive aspect) and on theory and prior research (the deductive aspect). Likewise, the working definition of the outcome to be explained, while open to revision as the research proceeds (the inductive aspect) is initially based on the researcher's preexisting knowledge and interests (the deductive aspect). It is appropriately labeled "induction," however, primarily because a core principle of AI is that researchers should attend to, and try to reconcile, disconfirming cases—*those that display the outcome in question, but not the hypothesized antecedent conditions*. Recall, from chapter 1, classic AI's pivotal fifth step: "If cases are found that challenge a working hypothesis, then either the antecedent conditions or the outcome must be reformulated in some way. Typically, if the outcome is reformulated, its scope is narrowed so that the nonconforming cases are excluded from the purview of the working hypothesis. If the antecedent conditions are reformulated, the causal argument is altered so that the nonconforming cases are embraced in some way, typically through a strategy of conceptual realignment. In either situation, the process of reformulation should be both public and transparent."

Consider a simple example: a researcher interested in terrorism examines causally relevant biographical details shared by a set of individuals who committed terrorist acts. The basic insight of AI is that the goal of understanding how outcomes happen dictates a focus on the causally relevant commonalities shared by positive instances. Studying negative cases (e.g., non-terrorists) provides little or no insight regarding how things come about (e.g., acts of terrorism). Suppose, in this example, that the researcher assesses religious radicalization as a causally relevant commonality. Classic AI emphasizes the identification and evaluation of antecedent conditions that are uniform (i.e., invariant) across instances of an outcome. For example, if the researcher was able to identify terrorists who failed to display religious radicalization, then she would take this refutation of religious radicalization as a serious challenge to its importance as a causally relevant antecedent, and not simply treat the disconfirming cases as flukes worthy of demotion to the error vector. This focus on invariant connections provides researchers the motivation to progressively revise or refine their working hypotheses, as they simultaneously deepen their knowledge of their cases.

Suppose further that the disconfirming cases (terrorists who did not display the antecedent condition, religious radicalization) nevertheless experienced a process of secular ideological radicalization. The researcher might decide to realign her working hypothesis to accommodate the new evidence, positing "religious *or* secular ideological radicalization" as a shared antecedent condition. Enlarging the scope of antecedent conditions is one of the key reformulation strategies addressed in this chapter.

As an alternative to reformulating explanatory concepts, a researcher might choose instead to accommodate disconfirming cases by narrowing the scope of the outcome. The goal of this strategy is to exclude disconfirming cases from the analysis altogether. This exact tactic was used by Howard Becker (1953, 2015) in the AI classic *Becoming a Marihuana User*. As we saw in chapter 1, Becker found that most users traveled through a series of steps in the process of achieving the outcome—becoming a user. However, he also discovered that some users did not go through the standard set of steps (Becker 1998: 205). Further research solved the puzzle: users who went through the steps learned to use marijuana *for pleasure*, while those who did not go through the steps did not learn to get high and used marijuana simply to appear to be "cool." By restricting his argument to those who used marijuana for pleasure, Becker was able to exclude the disconfirming cases from the purview of his working hypothesis.

ADDRESSING DISCONFIRMING CASES: A FORMALIZATION

The literature on AI emphasizes two main strategies, introduced above, for dealing with disconfirming cases. It is important to recognize, however, that there are several variants of each general strategy. The choice of strategies is based primarily on the researcher's close inspection of the disconfirming cases and careful comparison of disconfirming cases with consistent cases.

As a backdrop for the discussion of strategies, consider table 2-1, which tabulates an outcome against a causally relevant antecedent condition. The outcome in this hypothetical example is presence/absence of mass protest against the

TABLE 2-1 Initial findings

	No severe austerity	Severe austerity
IMF protest	Cell *a:* disconfirming cases; $N = 5$	Cell *b:* consistent cases; $N = 25$
Negligible or no IMF protest	Cell *c:* alternate-outcome cases; $N = 15$	Cell *d:* alternate-outcome cases; $N = 15$

$N = 60$.

International Monetary Fund (IMF); the causal condition is presence/absence of the imposition of severe austerity measures, mandated by the IMF as conditions for debt restructuring. The analysis embraces sixty less developed, debtor countries.[1] AI focuses on the first row of the 2 × 2 table, where the outcome is present. The researcher's goal is to identify antecedent conditions that are shared by instances of the outcome, as indicated by an empty cell *a* and a well-populated cell *b*.[2] In this example, twenty-five of the thirty instances of the outcome share severe austerity as an antecedent condition—five cases short of perfect consistency.

Cases in cell *a* are treated as anomalies to be addressed by the investigator. These cases gain specificity in the course of the research, as the researcher resolves inconsistencies through close inspection of the evidence and systematic comparison of cell *a* cases with cell *b* cases.[3] The primary focus is on strategies for emptying cell *a* of cases. There are two possible destinations for cell *a* cases: they can be moved to cell *b* or *c*. To move cell *a* cases to cell *b*, the researcher must reformulate the antecedent condition so that it is more *inclusive*. To move cell *a* cases to cell *c*, the researcher must reformulate the outcome so that it is more *restrictive*. There are two main variants of each strategy.

EXPANDING THE SCOPE OF THE ANTECEDENT CONDITION

The first variant of this strategy involves using the logical term *or* to join two (or more) related antecedent conditions, as in the terrorist radicalization example discussed above. In the context of the present example, IMF protest, assume that the researcher examined cell *a* cases and concluded that even though these cases were not subjected to severe austerity measures, there was still substantial IMF protest due to each country's heavy debt burden. The researcher combines these two conditions as shown in table 2-2, which illustrates the impact of treating "heavy debt burden" and "severe austerity measures" as substitutable antecedent conditions. The use of logical *or* to join two or more conditions entails a reconceptualization of the two conditions as a single, more abstract condition. In this example, the antecedent condition might be reformulated as "debt induced economic hardship."

TABLE 2-2 Using logical *or* to increase the scope of an antecedent condition*

	No severe austerity and no heavy debt burden	Severe austerity or heavy debt burden
IMF protest	Cell *a:* disconfirming cases; $N = 0$	Cell *b:* consistent cases; $N = 30$
Negligible or no IMF protest	Cell *c:* alternate-outcome cases; $N = 12$	Cell *d:* alternate-outcome cases; $N = 18$

*Compare with table 2-1.

TABLE 2-3 Lowering the threshold of the antecedent condition*

	Less than moderate austerity	Moderate to severe austerity
IMF protest	Cell *a:* disconfirming cases; $N = 0$	Cell *b*: consistent cases; $N = 30$
Negligible or no IMF protest	Cell *c:* alternate outcome cases; $N = 11$	Cell *d*: alternate outcome cases; $N = 19$

*Compare with table 2-1.

Comparing table 2-2 to table 2-1, the five cell *a* cases have moved to cell *b*, effectively emptying cell *a* of cases and establishing a pattern of results consistent with the goals of AI. Note that this reformulation of the antecedent condition also moves three cell *c* cases to cell *d*. Thus, from the viewpoint of a statistical assessment, there is only modest gain; however, from the viewpoint of AI, the researcher has successfully reformulated the working hypothesis and established an invariant connection.

The second variant of this strategy focuses on thresholds (table 2-3). The controlling principle is that the researcher assesses the degree to which the antecedent condition must be present for the outcome to be triggered. Assume that the researcher examined cell *a* cases and concluded that even though these cases were not subjected to severe austerity, they nevertheless experienced substantial IMF pressure in the form of moderate austerity measures. This pattern of results suggests that the initial threshold for the antecedent condition, severe austerity, was too high. Only moderate austerity was required. The researcher decides to relabel and recalibrate the antecedent condition, consistent with the discovery that moderate austerity engenders IMF protest. The five cell *a* cases relocate to cell *b*, thereby establishing a pattern of results consistent with the goals of AI. Note that several cases also shift from cell *c* to cell *d*, so once again there is only modest gain from a statistical viewpoint, but from the viewpoint of AI the reformulation is decisive—cell *a* is empty and an invariant connection has been established.

Restricting the Scope of the Outcome

The second general strategy involves narrowing the set of cases with the outcome, in an effort to relocate cell *a* cases to cell *c* in table 2-1. Suppose that, following close inspection of cell *a* cases and the comparison of cell *a* cases with cell *b* cases, the

TABLE 2-4 Qualifying the outcome, making it more restrictive*

	No severe austerity	Severe austerity
Broad-based IMF protest	Cell *a:* disconfirming cases; $N = 0$	Cell *b:* consistent cases; $N = 22$
Negligible or no broad-based IMF protest	Cell *c:* alternate-outcome cases; $N = 20$	Cell *d:* alternate-outcome cases; $N = 18$

*Compare with table 2-1.

TABLE 2-5 Raising the outcome threshold*

	No severe austerity	Severe austerity
Acute IMF protest	Cell *a:* disconfirming cases; $N = 0$	Cell *b:* consistent cases; $N = 19$
Non-acute or no IMF protest	Cell *c:* alternate-outcome cases; $N = 20$	Cell *d:* alternate-outcome cases; $N = 21$

*Compare with table 2-1.

researcher observes that in contrast to most cell *b* cases, the protest in cell *a* cases was not broad based. Instead, it was driven mostly by labor unions. This difference between cell *a* cases and most cell *b* cases provides an opportunity to reformulate the outcome in a way that excludes cell *a* cases.

Table 2-4 shows the impact of reformulating the outcome. It has been changed from "IMF protest" to "broad-based IMF protest," and cases have been shifted to accommodate the reformulation. The five cell *a* cases now reside in cell *c*, and several cell *b* cases shift to cell *d* because they did not meet the revised outcome standard (broad-based IMF protest). Once again, from a statistical standpoint, there has been only modest gain; but, from the perspective of AI, there is now the clarity of an invariant connection between the causal condition and the outcome.

The second variant of the outcome-based strategy focuses on thresholds. In this instance, suppose that the researcher compares cell *a* cases with cell *b* cases and concludes that most cell *b* cases had widespread, violent protest against the IMF, while no cell *a* cases had such acute levels. She decides to raise the threshold for the outcome to "acute" IMF protest and reassigns cases to cells accordingly. The five disconfirming cases relocate to cell *c*, and several cell *b* cases are reassigned to cell *d*. Table 2-5 illustrates the impact of raising the outcome threshold. Consistent with the goals of AI, an invariant connection has been established, and once again, a reconciliation strategy that advances understanding from the perspective of AI registers only modest gain from a statistical viewpoint.

TABLE 2-6 Imposing a scope condition: low-income countries*

	No severe austerity	Severe austerity
IMF protest	Cell *a:* disconfirming cases; $N = 0$	Cell *b:* consistent cases; $N = 20$
Negligible or no IMF protest	Cell *c:* alternate-outcome cases; $N = 12$	Cell *d:* alternate-outcome cases; $N = 13$

*$N = 45$; compare with table 2-1.

TWO MORE STRATEGIES

While most treatments of AI emphasize reformulating the antecedent conditions or the outcome, as illustrated in tables 2-2 through 2-5, two additional strategies warrant attention: (1) stipulating a scope condition and (2) typologizing the outcome.

Stipulating a Scope Condition

The first strategy is to specify a "scope condition" that can be used as a filter to restrict the set of relevant cases (Walker and Cohen 1985; Goertz and Mahoney 2012: 205–17). For example, suppose that a researcher observes that the five cases in cell *a* of table 2-1 are all middle-income countries, while most cell *b* countries are low-income countries. The researcher speculates that the close connection between severe austerity and IMF protest may be specific to low-income countries. The researcher decides to use "low-income countries" as a scope condition and removes all middle-income countries from the analysis. The results are shown in table 2-6.

The results reveal an invariant connection between severe austerity and IMF protest, specific to low-income countries. While this pattern of results satisfies the requirements of AI, from the perspective of statistical analysis it does so at a substantial cost. Observe the number of cases in table 2-6 versus table 2-1. Stipulating a scope condition reduces the sample size from $N = 60$ to $N = 45$. Tests of statistical significance are very powerfully influenced by the number of cases—the fewer the cases, the more difficult it is to achieve significance. Thus, using a scope condition to narrow the set of relevant cases may jeopardize statistical significance. Dropping middle-income countries in this example impacts the count of cases in all four cells. Thus, from the perspective of statistical analysis, table 2-6 offers only slight gain, at best, over table 2-1, despite the fact that greater empirical clarity has been achieved via AI.

Specifying Subtypes of an Outcome

The final analytic strategy for dealing with disconfirming cases involves distinguishing subtypes of the focal outcome. This strategy is not part of the corpus of classic AI, but instead is a logical extension of basic principles of the approach, relevant to generalized AI (the focus of part II of this book). Like the third and fourth strategies discussed above, specifying subtypes involves reformulating the outcome as a way to cope with disconfirming cases (George and Bennett 2005).

As before, the focus is on cell *a*—cases that display the outcome but not the hypothesized causal condition(s). The researcher asks: Did cell *a* cases experience an outcome that differed qualitatively from the outcome experienced by cell *b* cases? If so, then the evidence may be reformulated in terms of outcome subtypes, with different causal conditions linked to different subtypes. From this viewpoint, the five cases in cell *a* of table 2-1 are not disconfirming, per se, but instead constitute the starting point of a separate application of AI, with an outcome that differs in kind from the outcome experienced by most cell *b* cases. Thus, the evidence in table 2-1 would prompt a reconceptualization of the outcome in terms of subtypes, and instead of simply demoting cell *a* cases to cell *c*, they would be made the focus of a separate analysis.

This strategy is similar, in some respects, to the outcome-oriented strategy depicted in table 2-4. In that example, cell *a* cases were reassigned to cell *c* because the IMF protests they exhibited were not broad based (the reformulated outcome). The same observation—that the protests exhibited by cell *a* cases were union based, while the protests exhibited by most cell *b* cases were broad based—is used in the present strategy to distinguish subtypes of IMF protest. The researcher's next step would be to remove cell *a* cases from table 2-1 (along with any cell *b* cases that were union based) and assess the antecedent conditions they share, in a completely separate application of AI. In the wake of their departures, these cases would leave behind an empty cell *a* and also a consistent antecedent condition (severe austerity) for a specific subtype of IMF protest: broad based.

It is important to note that the sixth strategy—typologizing the outcome—differs fundamentally from the practice of positing equifinality (Mackie 1965, 1980). To allow for equifinality (a core feature of qualitative comparative analysis) is to acknowledge that there may be different causal recipes for the same outcome. From the viewpoint of equifinality, cases in cell *a* of table 2-1 are simply cases that experienced a different, but causally equivalent, recipe for the outcome in question (IMF protest). It is the researcher's task to identify alternate but equivalent causal recipes. By contrast, the typologizing strategy accepts, in principle, that cases residing in cell *a* exhibit a different causal recipe, but adds the stipulation that the researcher should identify differences in the outcome that follow from differences in causal recipes. For example, union-based IMF protest might be limited to strikes and peaceful demonstrations, while broad-based IMF protest might include additional, more violent forms of protest (e.g., riots).

EXTENSIONS

The examples offered so far use simple 2 × 2 tables with only a single causal condition in each illustration (except for table 2-2, which joined two causal conditions using logical *or*). In practice, researchers are more likely to identify multiple antecedent conditions shared by instances of an outcome. When attempting to account for "how things happen," it is useful to think in terms of causal

TABLE 2-7 Assessing a causal recipe

	Causal recipe not satisfied	Causal recipe satisfied*
IMF protest	Cell *a:* disconfirming cases	Cell *b:* consistent cases
Negligible or no IMF protest	Cell *c:* alternate-outcome cases	Cell *d:* alternate-outcome cases

*Causal recipe = severe austerity combined with government corruption, prior mobilization, and high inflation.

recipes—combinations of conditions joined by logical *and*—that generate the outcome in question. Thus, rather than cross-tabulating single conditions against an outcome, the researcher would instead focus on the relevant antecedent conditions shared by instances of an outcome, and assess the consistency of causal *recipes*. As before, the focus is on cells *a* and *b* of the cross-tabulation, but the column headings would indicate the presence/absence of a causal recipe.

Table 2-7 offers a simple illustration. Again, the outcome is protest against the IMF. The causal recipe has four ingredients: severe austerity combined with government corruption, prior mobilization, and high inflation. Disconfirming cases (cell *a*) are those that display the outcome but not the causal recipe; consistent cases reside in cell *b*; cases satisfying the causal recipe but not displaying the outcome (i.e., "alternate-outcome" cases) are in cell *d*; while cases residing in cell *c* are alternate-outcome cases that failed to satisfy the causal recipe in question.

Another limitation of the examples offered so far is that they rely on present/absent causal conditions and present/absent outcomes. Social scientists often deal with phenomena that vary by level or degree. For example, inflation can be precisely measured. To dichotomize it as "high" versus "not-high" in a causal recipe (as in table 2-7) may seem wasteful of useful information. Fortunately, there is a readymade solution, which is to calibrate causal conditions and outcomes that vary by level or degree as fuzzy sets (Zadeh 1965, 1972; Kosko 1993; Ragin 2000, 2006a, 2008; Ragin and Fiss 2017; Smithson 1987; Smithson and Verkuilen 2006; see also appendix B). With fuzzy sets, it is possible to evaluate the *degree* of membership of each case in each relevant set. Membership scores range from 0 to 1, with a score of 0 indicating full non-membership, a score of 1 indicating full membership, and a score of 0.5 (the crossover point) indicating maximum ambiguity in whether a case is more in or more out of the set in question. For example, a country might be assigned a membership score of 0.80 in the outcome set, IMF protest, indicating that it has strong but not quite full membership in the outcome. The calibration of fuzzy-set membership scores is heavily knowledge dependent and should be based as much as possible on external criteria, and not on inductively generated criteria such as means and standard deviations or percentiles (Ragin 2000; Ragin 2008: chaps. 4 and 5).

For illustration, consider figure 2-1, a scatterplot showing a hypothetical relation between degree of membership in a causal recipe (severe austerity combined with government corruption, prior mobilization, and high inflation) and degree of

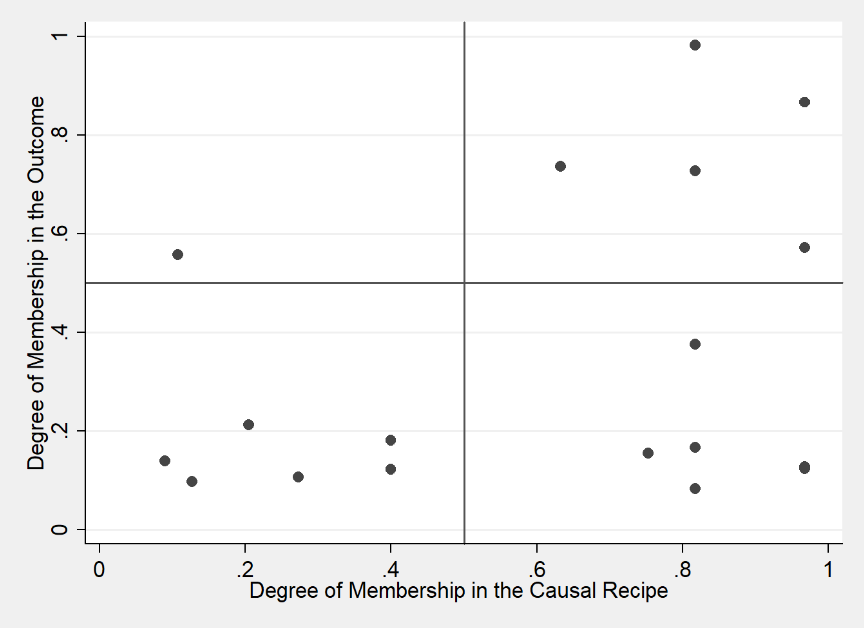


FIGURE 2-1. Illustration of the use of fuzzy-set membership scores.

membership in an outcome (IMF protest). Degree of membership in the causal recipe is calculated by first calibrating the four conditions as fuzzy sets and then selecting, for each case, the lowest of its four fuzzy membership scores, which becomes that case's degree of membership in the causal recipe. Using the lowest membership score directly implements fuzzy-set intersection, an operation that follows "weakest link" reasoning. A case can be assigned greater than 0.5 membership in a causal recipe only if it has greater than 0.5 membership in each component of the recipe.

The plot in figure 2-1 is divided into four quadrants using the two crossover points (scores of 0.5 on the causal recipe and on the outcome). The central focus of AI is the top half of the plot—cases that are more in than out of the set of cases with the outcome. Cases residing in the top-right quadrant are consistent cases—they share greater than 0.5 membership in the outcome and the causal recipe. Cases residing in the top-left quadrant are more in than out of the outcome set, but do not exhibit strong membership in the causal recipe. Thus, cases in this quadrant are disconfirming cases. It is the researcher's goal to reconcile these cases using the strategies described in this chapter. Cases residing in the lower-right quadrant share membership in the causal recipe but not in the outcome and are treated as "alternate-outcome" cases deserving of separate analytic attention (i.e., an assessment of "what happened instead"). Likewise, cases residing in the bottom-left quadrant are also alternate-outcome cases, and not directly relevant to AI.

DISCUSSION

The strategies for reconciling disconfirming cases described in this chapter involve close inspection of disconfirming cases and careful comparison of disconfirming cases with consistent cases. The choice of which reconciliation strategy to use is based fundamentally on the knowledge gained through case-oriented investigation. Because all strategies focus exclusively on emptying cell *a* of cases, they differ substantially from strategies rooted in conventional statistical methods. From the viewpoint of conventional statistical methods, the relationship observed in table 2-1 between severe austerity and IMF protest is simply probabilistic. There may be additional independent variables that could be added to the analysis that would increase the accuracy of prediction, but there is no specific focus on any one cell, nor is there any singular interest in establishing an invariant connection.

The strategies described in this chapter do not directly address the distribution of cases in the second row of the 2 × 2 table—where the outcome is absent. AI seeks to establish invariant connections between antecedent conditions and the presence of an outcome, paying relatively little parallel attention to the absence of the focal outcome. As explained in chapter 4, investigating the second row of table 2-1, especially cases in cell *d*, requires specification of "what happened instead?"—which could involve several alternate outcomes. Through the lens of AI, instances of the "absence" of an outcome are not "negative" cases; rather, they are positive cases of one or more alternate outcomes and are deserving of separate analytic treatment.

Finally, it is important to reiterate that the reconciliation strategies described in this chapter should be implemented in a completely transparent manner. These strategies entail close inspection of cell *a* cases and careful comparison of these cases with cases in cell *b*. The researcher has gained important insights from comparative case analysis, and she should indicate what she has learned and how she learned it.

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

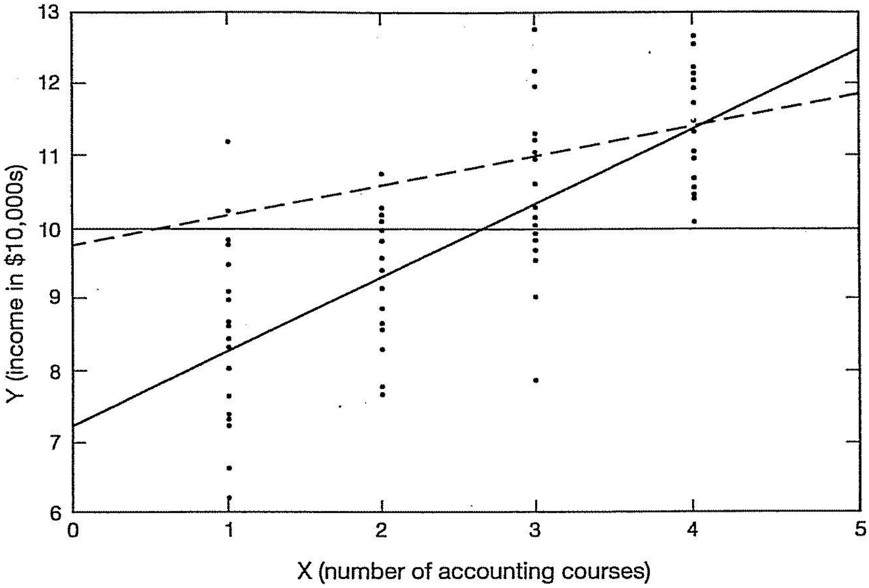


FIGURE 3-1. Recent MBA income levels plotted against number of accounting courses completed (from King et al. 1994: 131).

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.

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.

PART TWO

Generalized Analytic Induction

5

Classic versus Generalized Analytic Induction

Chapters 1–4 describe major features of analytic induction (AI), especially its distinctive logic. These features, in turn, provide the basis for the formulation of generalized AI. Not all features of classic AI are retained by generalized AI, however, and some readers may regard the omitted features as too important to be discarded. Still, much of the logic and spirit of classic AI is captured in the generalized version. This chapter presents a brief overview of the features that are retained and those that are left behind.

FOUNDATIONAL FEATURES OF GENERALIZED ANALYTIC INDUCTION

Importance of Addressing Research Questions Regarding Qualitative Outcomes

Both classic AI and generalized AI focus on qualitative outcomes that can be conceived as "happenings" or instances. The types of phenomena that can be viewed in this way are varied in nature, ranging from micro-level interactions to social revolutions. Even interval-scale dependent variables can be transformed into evidence on happenings. Using fuzzy sets, for example, data on household income can be transformed into an assessment of the degree of membership in the set of households that are at least "middle class," an achieved status. The key is that external knowledge (e.g., how much income is required to be considered middle class) is applied to quantitative evidence in order to establish a qualitative distinction. As explained in chapter 3, AI focuses on "how things happen," which, in turn, means that the outcome is typically qualitative in nature.

Importance of Centering the Analysis on Examination of Cases with the Focal Outcome and Their Shared Antecedent Conditions

Both classic AI and generalized AI focus more or less exclusively on positive cases—instances of the outcome in question. While most analytic techniques attend to the differences between positive and negative cases (i.e., those with the focal outcome versus those without it), AI examines cases with the outcome in question. Because a given set of cases may exhibit several different outcomes, AI focuses on one outcome at a time. AI identifies antecedent conditions for each outcome, which in turn provides the basis for the interpretation of each outcome's etiology.

Utilization of Diverse Strategies to Address and Reconcile Positive/Disconfirming Cases

Cases displaying the focal outcome but lacking the antecedent conditions specified in a working hypothesis provide important clues regarding how to (1) revise a working hypothesis or (2) reconceptualize the outcome. Positive/disconfirming cases can be reconciled in a variety of ways, depending on what is learned from comparing positive/disconfirming cases with positive/confirming cases. Both classic AI and generalized AI inductively refine working hypotheses based on careful consideration of positive/disconfirming cases.

Rejection of the Analysis of Heterogeneous Complements

Most conventional analytic techniques attempt to identify the conditions that influence the level, degree, or probability of an outcome, conceived as a dependent variable. These techniques all depend on the existence of variation in the outcome. When the outcome is presence/absence, the binary opposition between presence and absence is typically lopsided: cases on one side of the dichotomy (usually the "presence" side) are relatively homogeneous with respect to the focal outcome, while cases on the other side of the dichotomy are heterogeneous, united only by their failure to display the focal outcome. Neither classic AI nor generalized AI is analytically dependent on having variation in outcomes, and both thereby avoid the problem of heterogeneous complements.

Reconceptualization of "Negative" Cases as Instances of Alternate Outcomes, Worthy of Separate Analytic Treatment

From the perspective of both classic AI and generalized AI, it is often prudent to unpack a heterogeneous complement and differentiate the various alternate outcomes contained within it. Rather than viewing negative cases as a catchall category, classic AI and generalized AI view them as collections of alternate outcomes, each outcome potentially worthy of separate analysis and assessment. From the perspective of AI, there are no negative cases, per se, only positive cases of different outcomes.

FEATURES OF CLASSIC AI THAT THE GENERALIZED VERSION LEAVES BEHIND

Generalized AI parts company with several central features of classic AI. The features that are left behind are the most controversial, as reflected in the debates in the early 1950s regarding the place of "universals" in social research (see chapter 1). The controversial nature of these features of classic AI is also reflected in the various attempts to moderate the approach—for example, by renaming it "modified" AI (Gilgun 1995), "neo" AI (Hicks 1994), "analytic fieldwork" (Katz 1983), or "generalized" AI, as in the present effort. Becker (1998: 196) notes that "researchers seldom use Analytic Induction in its classical form," but "in slightly less rigorous and single-minded versions, it is widely used."

Determined Pursuit of Invariant Connections

An application of classic AI is not considered truly complete until the investigator establishes an invariant connection between one or more antecedent conditions and the outcome. Ideally, the match should be perfect, with no positive/disconfirming cases remaining. By contrast, generalized AI is satisfied by connections that are imperfect, but largely consistent. Of special interest to generalized AI are *modal configurations*—combinations of antecedent conditions that are relatively common in a set of cases exhibiting the outcome in question. Thus, generalized AI relaxes classic AI's demand for invariance, seeking instead to identify common clusters of antecedent conditions linked to the focal outcome.

Proactive Search for Disconfirming Cases

Classic AI demands that researchers doggedly seek out positive/disconfirming cases. After all, the goal of classic AI is to continue to refine the working hypothesis or the definition of the outcome until an invariant connection (i.e., a "universal" relationship) has been established. From the perspective of classic AI, the potential set of cases relevant to an investigation is unrestricted; the key consideration regarding case selection is whether a case challenges the working hypothesis. With generalized AI, by contrast, the set of relevant cases may be established in advance, and it is permissible to examine connections between antecedents and outcomes within a defined population, a sample, or any meaningfully circumscribed set of cases.

Rejection of Frequency Criteria

Classic AI has little interest in the use of frequency criteria. Znaniecki (1934) made this position clear in his assault on "enumerative" induction, arguing that AI was more scientific because of its emphasis on establishing "universal" relationships. By contrast, generalized AI permits imperfection. Once less-than-perfect connections are permitted, enumerative criteria regain importance. For example, it matters whether a combination of antecedent conditions—a modal configuration—is found in 60 percent or 85 percent of the cases exhibiting an outcome. More is better, and *more* is best established using enumerative criteria.

Focus on Singular Outcomes

When confronted with positive/disconfirming cases, one common classic AI response is to narrow the definition or the empirical scope of the outcome (e.g., shifting from a focus on "marijuana users" to a narrower focus on "users of marijuana for pleasure"). This reconciliation strategy seeks simply to exclude the disconfirming cases from an analysis. By contrast, generalized AI allows the specification of subtypes of the focal outcome, with different antecedent conditions linked to different subtypes. Thus, rather than excluding disconfirming cases altogether (e.g., excluding marijuana users, not for pleasure), generalized AI allows specification of subtypes of the outcome (e.g., for pleasure vs. not for pleasure), linked to differences in antecedent conditions.

6

The Interpretive Logic of Generalized Analytic Induction

In his study of addiction, Alfred Lindesmith (1968) focused exclusively on conditions that made sense as contributing causes, and searched for invariant connections between the outcome—addiction—and relevant antecedent conditions. He observed an important commonality shared by all opiate addicts: they succumbed to addiction after an explicit and abrupt *recognition* that a long-standing pattern of distress had been a result of repeated opiate withdrawal (Katz 2001) and not of some other ailment. Lindesmith did not treat *recognition* as a variable (i.e., as something that varied systematically across cases) because he was interested only in the consistency of its *presence* as an antecedent condition in instances of opiate addiction (see chapter 1).

Lindesmith's analytic strategy reflects AI's distinctive approach to the assessment of empirical evidence—specifically, how a data set on multiple cases is employed to generate results. In this regard, AI differs from both conventional quantitative analysis and qualitative comparative analysis (QCA; Ragin 1987). Both conventional quantitative analysis and QCA investigate causally relevant conditions that vary by level, degree, or presence/absence.[1] As this chapter demonstrates, generalized AI evaluates the two sides of a binary causal condition *not* as "present versus absent" but as "contributing versus irrelevant." In this approach to evidence, only one side of a binary is considered important; the other side is typically interpreted as "not contributing" and is excluded from consideration (Hammersley and Cooper 2012: 140).

For example, if "state breakdown" is considered a relevant contributing cause of social revolution (as in Skocpol 1979), then the absence of state breakdown can be eliminated from consideration as a possible contributing cause, across all cases included in the analysis. AI typically selects one side of a presence/absence dichotomy as relevant to an outcome, and treats the other side of the dichotomy as irrelevant (Hammersley and Cooper 2012: 155). The evaluation of each condition as contributing versus irrelevant is based on the researcher's substantive and theoretical knowledge, and thus involves interpretive inferences. This aspect of AI follows directly from its roots in qualitative research.

The main contrast addressed in this chapter is between QCA (Ragin 1987) and generalized AI. The contrast with QCA serves to highlight the distinctiveness of generalized AI. The discussion of the chasm separating generalized AI and conventional quantitative analysis is limited, for the simple reason that quantitative techniques require variation in both outcomes and causal conditions. The idea that a causal condition is either contributing or irrelevant is completely foreign to conventional quantitative analysis, which is wedded to the principle of covariation, which in turn requires variation in both antecedent conditions and outcomes. Generalized AI requires neither. For example, if positive instances of social revolution all exhibit state breakdown as an antecedent condition, then neither the antecedent condition nor the outcome varies across relevant cases.

QCA AND POSITIVE CASES

Most QCA applications include both positive and negative instances of the outcome in question. These values, in turn, shape the coding of truth table rows as "true" (causal combinations linked to outcome) or "false" (combinations not linked to outcome). Truth table rows that cannot be coded "true" or "false" on the outcome (typically due to a lack of cases) are called "remainder" rows. Researchers use the remainder rows to craft truth table solutions that are simpler than the "complex" solution (for a discussion of complex, parsimonious, and intermediate solutions, see Ragin 2008: chap. 9). Thus, the typical truth table analysis has three types of rows: true, false, and remainder. The remainder category embraces all truth table rows that cannot be coded true or false.

It is not generally recognized that QCA is capable of analyzing a body of evidence that contains only positive instances of an outcome. When used in this manner, QCA codes truth table rows "true" if they contain instances of the outcome, while rows that are devoid of cases are classified as remainder rows. Thus, in this type of application, there are only two kinds of truth table rows: true (contains instances of the outcome) and remainder (no instances).[2] However, with this setup, the remainder rows cannot be used to craft simpler solutions (i.e., intermediate and parsimonious). Remainder rows are incorporated into truth table solutions if doing so produces a logically simpler solution. However, the results in this setup, with only two kinds of truth table rows, are degenerate because *all* logically possible combinations of conditions (positive and remainder) can be linked to the outcome in question, which is not a meaningful truth table solution. Instead, all remainder rows must be treated as false. The upshot: if an application has only positive cases, the parsimonious and intermediate solutions to the truth table cannot be derived. Only the complex truth table solution is possible.[3]

Nor is it generally recognized that QCA's "complex" solution to truth table analysis uses only truth table rows with coded outcomes equal to 1 (true), even in applications where there are both positive and negative instances of the outcome and thus all three kinds of truth table rows—true, false, and remainder. To generate the complex solution, truth table rows with outcome equal to 1 are paired and compared with each other, in an attempt to eliminate conditions one at a time through a "bottom-up" process known as incremental elimination. Consider, for example, an analysis with four causal conditions (A, B, C, and D) and an outcome (Y). If A•~B•C•D → Y and A•~B•~C•D → Y, it is possible to eliminate condition C/~C when conditions A, ~B, and D are present, yielding A•~B•D → Y (tilde indicates negation or *not*; arrow indicates the superset/subset relationship; multiplication symbol indicates combined conditions; plus sign indicates alternate combinations or alternate conditions). Condition C/~C is eliminated in this particular context (A•~B•D), but not in other contexts (e.g., A•B•D). To eliminate two conditions, four rows, all coded 1 (true) on the outcome, must be matched. For example, if A•B•C•D, A•B•~C•D, A•~B•C•D, and A•~B•~C•D are all coded 1 (true) on the outcome, then both B/~B and C/~C can be eliminated, yielding A•D → Y.[4] To eliminate three conditions, eight rows with the outcome must be matched, and so on. These requirements follow directly from QCA's configurational logic.

For QCA's complex solution to yield useful results, it is important to have a nontrivial proportion of truth table rows coded 1 (true). Consider, for illustration, Olav Stokke's (2004) truth table for successful shaming of violators of international fishing agreements (table 6-1). Please note that only Stokke's positive cases are shown in the truth table, which is all that is required to derive the complex solution. Using the three-letter condition labels, as shown in the table, the four truth table rows with positive cases can be rewritten as follows:

adv•~com•shd•inc•rev + adv•com•shd•inc•rev +
adv•com•shd•~inc•~rev +
adv•~com•~shd•~inc•~rev → success

The truth table rule for combining rows to reduce complexity is that two rows can be combined to create a simpler expression if they agree on the outcome (e.g., they are both coded "true") and differ on only one condition. This rule is clearly satisfied by the first two rows because they differ on only com/~com:

adv•~com•shd•inc•rev + adv•com•shd•inc•rev
= adv•shd•inc•rev •(com + ~com)
= adv•shd•inc•rev

TABLE 6-1 Stokke's truth table for successful shaming of violators (positive cases)

Advice (adv)	Commitment (com)	Shadow (shd)	Inconvenience (inc)	Reverberation (rev)	Success
1	0	1	1	1	1
1	1	1	1	1	1
1	1	1	0	0	1
1	0	0	0	0	1

NOTES:

Advice (adv): Whether the shamers can substantiate their criticism with reference to explicit recommendations of the regime's scientific advisory body.

Commitment (com): Whether the target behavior explicitly violates a conservation measure adopted by the regime's decision-making body.

Shadow of the future (shd): Perceived need of the target of shaming to strike new deals under the regime—such beneficial deals are likely to be jeopardized if criticism is ignored.

Inconvenience (inc): The inconvenience (to the target of shaming) of the behavioral change that the shamers are trying to prompt.

Reverberation (rev): The domestic political costs to the target of shaming for not complying (i.e., for being scandalized as a culprit).

However, this simplification is all that is possible for truth table 6-1, yielding the following complex solution:

$$\text{adv}\bullet\text{shd}\bullet\text{inc}\bullet\text{rev} + \text{adv}\bullet\text{com}\bullet\text{shd}\bullet{\sim}\text{inc}\bullet{\sim}\text{rev} + \text{adv}\bullet{\sim}\text{com}\bullet{\sim}\text{shd}\bullet{\sim}\text{inc}\bullet{\sim}\text{rev} \rightarrow \text{success}$$

In other words, because the diversity of positive cases is empirically limited in this example (with only four of the thirty-two logically possible combinations displaying the outcome), very little reduction of complexity can be realized.

In part, QCA's goal of reducing complexity is stymied, in this example, by one of its core strengths: its strict adherence to configurational logic. QCA gives equal analytic weight to the presence of conditions and to the absence of conditions. Consider, for example, the last row of the truth table: adv•~com•~shd•~inc•~rev. Three of the conditions that are combined in this expression (the *absence* of an "explicit commitment," the *absence* of a "shadow of the future," and the *absence* of "domestic reverberations") are thought to undermine the success of shaming when they are coded present, and not when they are coded absent. Yet with QCA the truth table is analytically open to the possibility that these three are required to be absent, and are essential to the success of shaming when combined with adv•~inc. QCA users routinely circumvent this limitation by deriving parsimonious and intermediate solutions. However, as noted previously, these two solution types are not derivable using QCA if there are only positive instances of the outcome. Lacking negative cases, and by implication lacking remainders as well, only the complex solution is derivable.

GENERALIZED AI AND POSITIVE CASES

As discussed above, generalized AI interprets conditions as either "contributing" (to the occurrence of the outcome) or "irrelevant." This view of causal conditions contrasts sharply with QCA's view. Using QCA, a condition becomes irrelevant only if it is linked to the outcome when the condition is present and when it is absent, across matched rows. Again:

$$A\bullet{\sim}B\bullet C\bullet D + A\bullet{\sim}B\bullet{\sim}C\bullet D \rightarrow Y$$
$$A\bullet{\sim}B\bullet D\bullet(C + {\sim}C) \rightarrow Y$$
$$A\bullet{\sim}B\bullet D \rightarrow Y$$

C/~C is demonstrably irrelevant, but only in the context of A•~B•D. C/~C could still be relevant in other contexts. This context-specific elimination of causal conditions follows directly from QCA's grounding in configurational logic.

Generalized AI, by contrast, offers a contrasting view and a different treatment of the same evidence (A•~B•C•D + A•~B•~C•D → Y). The foundation of generalized AI's interpretive logic is the researcher's knowledge and understanding of the connection between the causal conditions and the outcome in question. Essentially, the researcher specifies, for each causal condition, whether it contributes to the outcome when it is present or when it is absent.[5] For example, if condition C contributes to the outcome only when it is present (C), then it is irrelevant when it is absent (~C). If a case (or a truth table row) includes ~C (the absence of C) as a condition, then the condition can be dropped from the combination because it is irrelevant (i.e., non-contributing). Consider generalized AI's approach to the evidence used to illustrate QCA: A•~B•C•D + A•~B•~C•D → Y. Assume that the researcher interprets each of the four conditions as contributing when present, and otherwise as irrelevant. Combination A•~B•C•D becomes A•C•D, and combination A•~B•~C•D becomes A•D. Logically, A•C•D is included in (i.e., is a subset of) A•D, which leaves A•D → Y as the solution of A•~B•C•D + A•~B•~C•D → Y. Thus, the generalized AI solution is far simpler than QCA's solution of the same evidence. The difference follows directly from the application of generalized AI's interpretive logic versus QCA's configurational logic.

This same interpretive logic can be applied to Stokke's data in table 6-1. Assume that the researcher interprets conditions adv (advice), com (commitment), shd (shadow of the future), and rev (domestic reverberations) as contributing to the outcome (successful shaming) when present, and otherwise as irrelevant; and interprets condition inc (inconvenient) as contributing to the outcome when negated (~inc), and otherwise as irrelevant. The four truth table rows from table 6-1 are transformed by this interpretive logic as shown in table 6-2, which uses dashes to indicate irrelevant (i.e., non-contributing) conditions. Thus:

adv•~com•shd•inc•rev becomes adv•shd•rev
adv•com•shd•inc•rev becomes adv•com•shd•rev
adv•com•shd•~inc•~rev becomes adv•com•shd•~inc
adv•~com•~shd•~inc•~rev becomes adv•~inc

TABLE 6-2 Stokke's truth table for positive cases viewed through the lens of generalized AI*

Advice (adv)	Commitment (com)	Shadow (shd)	Inconvenience (inc)	Reverberation (rev)	Success
1	–	1	–	1	1
1	1	1	–	1	1
1	1	1	0	–	1
1	–	–	0	–	1

* Dashes replace non-contributing conditions.

Generalized AI's use of interpretive inferences, just demonstrated, is strongly rooted in the case-oriented logic of qualitative research. For example, consider how a qualitative researcher would assess the first combination listed above (adv•~com•shd•inc•rev) as a single case. Armed with the knowledge that shaming succeeded in this case, the researcher would examine its array of conditions and pinpoint those that contributed to the outcome. In this light, three conditions (adv•shd•rev) make sense as components of a recipe for the outcome; the other two (~com•inc) do not. This same interpretive logic applies, as well, to the other three truth table rows, considered as cases. When explaining each case, a qualitative researcher would construct a case narrative based on *contributing* conditions.

Further simplification of table 6-2 is possible using the inclusion rule, which allows more complex terms (subsets) to be absorbed by less complex terms (supersets):

adv•com•shd•rev is included in adv•shd•rev
adv•com•shd•~inc is included in adv•~inc

Thus, generalized AI's solution of the truth table is straightforward, especially when compared to QCA's complex solution. It is simply

adv•shd•rev + adv•~inc → success

According to generalized AI, there are two causal recipes for successful shaming: (1) supportive scientific advice (adv) in situations where it is not inconvenient for the target of shaming to alter its behavior (~inc), and (2) supportive scientific advice (adv) in situations where there are both domestic reverberations for being shamed (rev) and a need to strike future deals (shd).

GENERALIZED AI AND OUTCOME SUBTYPES

As noted previously, generalized AI focuses on causally relevant conditions shared by positive cases. The only universally shared condition in the example presented above is supportive scientific advice (adv). When viewed from a classic AI perspective, the other conditions (shd, rev, and ~inc) can be seen as disconfirming, because there are instances of the outcome lacking each one of these conditions (e.g., rows 1 and 2 both lack ~inc). However, recall that one of the key strategies discussed in chapter 2 for dealing with disconfirming cases is to differentiate subtypes of the outcome in accordance with the different causal recipes. In this example, the investigator would look for qualitative differences between instances of successful shaming generated by adv•~inc versus those generated by adv•shd•rev, and construct a simple, two-category typology of outcomes based on the key differences identified. The contrast would attend to outcome differences between cases in the first two truth table rows (instances of adv•shd•rev) versus cases in the third and fourth rows (instances of adv•~inc). In this example, the researcher might distinguish between successful shaming where compliance is "pro forma" (adv•~inc) and successful shaming where compliance is "strategic" (adv•shd•rev).

Notice also that there is logical overlap between the two recipes: instances of adv•~inc•shd•rev, if they existed, would conform to both recipes. It is possible to assign this overlap to recipe adv•~inc, and thereby clarify and separate the two causal recipes. The first step is to use De Morgan's theorem to derive the complement (negation) of the recipe selected to receive the overlap. Next, the complement (negation) of that recipe is intersected with the other recipe, which narrows the breadth of the second recipe while awarding the overlap to the first:

adv•~inc + adv•shd•rev	generalized AI solution
adv•~inc	selected to receive overlap
~(adv•~inc) = ~adv + inc	recipe negated
(~adv + inc)•adv•shd•rev	intersected with other recipe
adv•inc•shd•rev	results of intersection
adv•~inc + adv•inc•shd•rev	clarified AI solution

The clarified recipes reveal the importance of whether the behavioral change is inconvenient to the targets of shaming. If it is not inconvenient (~inc), then the conditions for successful shaming are simple, namely, supportive scientific advice (adv). However, if the behavioral change is inconvenient (inc), then two additional conditions for successful shaming require satisfaction, the need to strike future deals (shd) and domestic reverberations (rev).

The contrast between QCA's and AI's approaches to the analysis of positive-only cases, just sketched, is sharp. QCA is stymied by the limited diversity of cases and its strict adherence to configurational logic; generalized AI is liberated from these constraints by its use of interpretive inferences. While QCA can be used to generate simpler truth table solutions when analyzing evidence that embraces both positive and negative cases, from the perspective of generalized AI, "negative cases," per se, don't exist. They are simply cases that exhibit outcomes that are different from the focal outcome.

TABLE 6-3 Stokke's truth table for unsuccessful shaming of violators (negative cases)

Advice (adv)	Commitment (com)	Shadow (shd)	Inconvenience (inc)	Reverberation (rev)	Success
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	0	0

TABLE 6-4 Stokke's truth table for "negative" cases viewed through the lens of generalized AI

Advice (adv)	Commitment (com)	Shadow (shd)	Inconvenience (inc)	Reverberation (rev)	Success
–	0	0	1	0	0
–	0	0	1	–	0
0	0	0	1	0	0
–	–	–	1	0	0

WHAT HAPPENED INSTEAD?

As explained in chapter 4, rather than defining cases that lack the focal outcome as "negative cases," AI considers such cases as instances of different outcomes and therefore as deserving of separate treatment. The researcher first identifies noteworthy outcomes among the nonfocal cases. Next, the researcher ascertains the antecedent conditions relevant to each alternate outcome. The relevant antecedent conditions for the alternate outcomes may differ substantially from the ones linked to the focal outcome.

Stokke's study of shaming as a way to induce violators of international agreements to mend their ways includes "negative" cases (where shaming did not have the desired impact). It would be ideal to know what happened in each case, for there may be several different outcomes among the cases that did not respond positively to shaming. Nevertheless, Stokke's negative cases can be used to illustrate generalized AI's approach to the analysis of a set of cases lacking the focal outcome. This illustration assumes (1) that their outcomes—resistance—are relatively homogeneous and (2) that the relevant causal conditions are the reverse of the conditions linked to the focal outcome. In essence, Stokke's "negative" cases of successful shaming are transformed into positive cases of resistance and subjected to the same analytic procedures applied to Stokke's positive cases.

Table 6-3 presents Stokke's negative cases (shaming failed). There are four truth table rows coded 0 (false) with respect to the success of shaming. As mentioned above, the causal conditions used in this example are the same as those used in the analysis of the positive cases (see table 6-1). However, the interpretive inferences are now the reverse of those implemented in table 6-2. The researcher interprets conditions adv (advice), com (commitment), shd (shadow of the future), and rev (reverberations) as contributing to the outcome (shaming failed) when *absent*, and otherwise as irrelevant; and interprets condition inc (inconvenient) as contributing when *present* (inc), and otherwise as irrelevant. The four truth table rows from table 6-3 are transformed by this interpretive logic, as depicted in table 6-4, which uses dashes to indicate irrelevant (i.e., non-contributing) conditions.

Converting table 6-4 into equation form yields

~com•~shd•inc•~rev + ~com•~shd•inc +
~adv•~com•~shd•inc•~rev + inc•~rev → ~success

Once again, further simplification is possible using the inclusion rule, which allows more complex terms (subsets) to be absorbed by less complex terms (supersets):

~com•~shd•inc•~rev	is included in both	~com•~shd•inc and inc•~rev
~adv•~com•~shd•inc•~rev	is included in both	~com•~shd•inc and inc•~rev

Thus, the generalized AI solution of truth table 6-4 is straightforward:

~com•~shd•inc + inc•~rev → ~success

In other words, shaming fails when it is inconvenient for the target to conform and there are no domestic reverberations, or when such inconvenience is combined with no explicit violation of a commitment and no need to strike future deals.

It is instructive to clarify the two recipes by assigning their overlap (~com•~shd•inc •~rev) to one of the two recipes:

~com•~shd•inc + inc•~rev	generalized AI solution
inc•~rev	selected to receive overlap
~inc + rev	recipe negated
(~inc + rev)•(~com•~shd•inc)	intersected with other recipe
~com•~shd•inc•rev	results of intersection
~com•~shd•inc•rev + inc•~rev	clarified solution

The clarified solution shows the pivotal impact of domestic reverberations. When domestic reverberations are absent, shaming will fail if it is inconvenient for the target to change its behavior. However, when domestic reverberations are present, the inconvenience of the change must be combined with an absence of an explicit commitment and no need to strike future deals.

CONTINGENT CONDITIONS

This chapter has emphasized generalized AI's use of interpretive inferences to transform "present versus absent" dichotomies to "contributing versus irrelevant" dichotomies. In many situations, however, a researcher will suspect that a condition is "contributing when present" in some contexts, while in other contexts it is "contributing when absent (i.e., negated)"—in short, that the valence of a contributing condition may be *contingent* on the other conditions involved. In these situations, the researcher has the option of treating such conditions as conventional presence/absence dichotomies, in order to ensure that their contrasting contributions are modeled correctly. Also, once the truth table solution is generated, it is possible to clarify the solution in a way that highlights the contrasting impact of the condition in question (see example in appendix C).

LOOKING AHEAD

Generalized AI's use of interpretive inference is one of the cornerstones of the approach. Applications of generalized AI presented in chapters 7–9 all use the binary opposition "contributing versus irrelevant" for most antecedent conditions, in place of configurational logic's "present versus absent." By focusing on outcomes one at a time and applying interpretive inferences, generalized AI is able to generate simplified representations of cross-case patterns in situations where the outcome is the same for all cases. Chapter 7 presents a step-by-step demonstration of generalized AI, focusing on a common qualitative research design—namely, situations where the researcher has a set of cases selected for study precisely because they all exhibit the same outcome. Chapter 8 provides an illustration of a generalized AI investigation of multiple outcomes, based on a reanalysis of data published in 2006 by Jocelyn Viterna on women's mobilization into the Salvadoran guerrilla army. Chapter 9 demonstrates the application of generalized AI to conventional quantitative data, using the Black female sample from the National Longitudinal Survey of Youth.

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

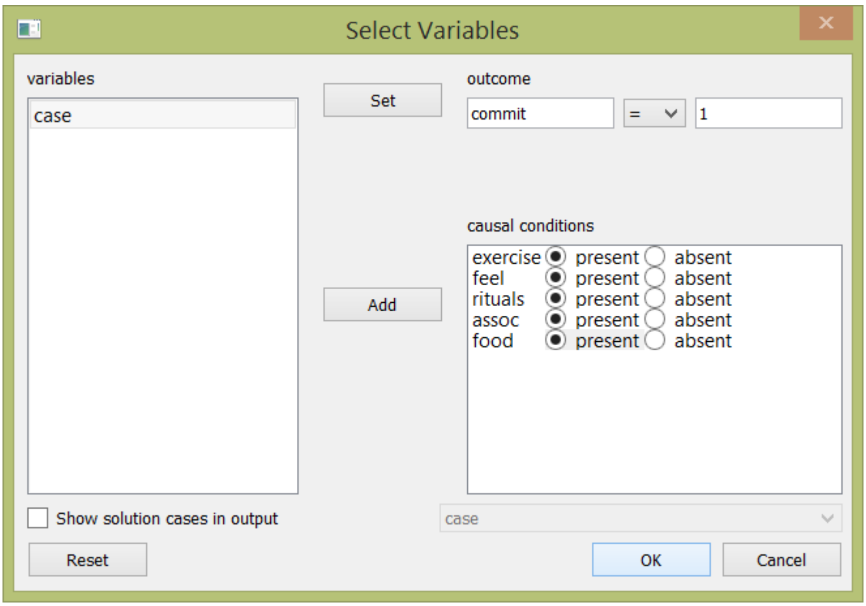


FIGURE 7-1. Dialogue box generating table 7-3.

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.

8

Using Generalized AI to Reanalyze Viterna's Study of Women's Mobilization into the Salvadoran Guerrilla Army

Jocelyn Viterna (2006) applies key principles of generalized AI in her pathbreaking study of women's mobilization as FMLN guerrillas in the Salvadoran Civil War (1980–92). Viterna distinguishes five different outcomes—three distinct paths to guerrilla activism (politicized, reluctant, and recruited) and two non-guerrilla paths (collaborators and nonparticipants). Rather than define the analysis as a binary contrast between the three guerrilla paths versus the two non-guerrilla paths, she focuses instead on the separate conditions linked to each of the five outcomes. In other words, she views each of the five outcomes as worthy of separate analytic attention and avoids conventional dichotomization of the outcome as "guerrilla versus non-guerrilla." This feature of her study aligns well with generalized AI as described in this book.

A respondent is categorized as a guerrilla if she "lived and worked in or alongside an FMLN guerrilla camp as a primary, permanent residence . . . for at least six months" (Viterna 2006: 16). The thirty-eight respondents classified as politicized, reluctant, or recruited guerrillas (the three paths to guerrilla activism) all met this criterion. *Politicized* guerrillas ($N = 7$) were motivated to join the guerrillas by their opposition to the Salvadoran government. *Reluctant* guerrillas ($N = 14$) joined the guerrillas following a crisis, and typically faced an absence of viable alternatives to joining the guerrilla camps. *Recruited* guerrillas ($N = 17$) were residents of refugee camps who were persuaded to join the guerrillas by members of the FMLN. Collaborators ($N = 12$), by contrast, "maintained a household as a primary residence, but held a formally defined role of support for the guerrilla camp." Finally, nonparticipants ($N = 32$) "maintained a household as a primary residence and did not hold any formal positions of support for the guerrillas" (2006: 16).

TABLE 8-1 Conditions relevant to guerrillas

Condition (name)	Description	Measurement
Previous organizational involvement (previnv)	Participated in a political or religious organization advocating reform (predating FMLN mobilization)	yes = 1, no = 0
Family ties (activefam)	Had FMLN family member(s), predating or simultaneous with FMLN mobilization	yes = 1, no = 0
Refugee/repopulated community (rcamp)	Lived in a refugee camp or repopulated community at moment of FMLN mobilization	yes = 1, no = 0
Motherhood (mother)	Had children at moment of FMLN mobilization	yes = 1, no = 0
Family completeness (complete)	Either with parents or partner at moment of FMLN mobilization	yes = 1, no = 0
Age (young)	Age at FMLN mobilization	7–17 years = 1, 18+ years = 0
Mobilization period (early)	FMLN mobilization occurred early or late	1980–83 = 1, 1985–91 = 0

TABLE 8-2 Conditions relevant to collaborators and nonparticipants

Condition (name)	Description	Measurement
Previous organizational involvement (previnv)	Participated in a political or religious organization advocating reform, prior to or during the war	yes = 1, no = 0
Family ties (activefam)	Had FMLN family member(s), prior to or during the war	yes = 1, no = 0
Refugee/repopulated community (rcamp)	Lived in a refugee camp or repopulated community at some point during the war	yes = 1, no = 0
Motherhood (mother)	Had children prior to or during the war	yes = 1, no = 0
Family completeness (complete)	Had either parents or partner during the entire length of the war	yes = 1, no = 0

Tables 8-1 and 8-2 list the background conditions Viterna used in her five analyses. The conditions differ for guerrillas and non-guerrillas. Not only are some conditions not relevant to non-guerrillas (e.g., regarding the timing of becoming a guerrilla), but there are also differences in contexts. For example, "motherhood" for guerrillas refers to the period prior to mobilization as guerrillas, while for non-guerrillas it refers to the condition of motherhood at any point prior to or during the civil war. More generally, it is important to point out that when researchers study multiple outcomes, it is not unusual for the relevant antecedent conditions to differ, sometimes substantially, across outcomes.

TABLE 8-3 Tabular data on politicized guerrillas

name	young	early	mother	complete	rcamp	previnv	activefam
Vilma	1	1	0	1	0	1	1
Alicia	1	1	0	1	0	1	1
Estela	1	1	0	1	0	1	1
Pati	0	1	1	1	0	1	1
Zoila	0	0	1	1	0	1	1
Gregoria	1	1	0	1	0	1	0
Gloria	1	1	0	0	0	1	0

TABLE 8-4 Data on politicized guerrillas converted to recipes

young	early	mother	previnv	activefam	*N*
1	1	0	1	1	3
1	1	0	1	–	2

POLITICIZED GUERRILLAS

Table 8-3 summarizes the tabular data that Viterna presents on the seven guerrillas she classifies as "politicized." She states that these guerrillas were pulled into participation by their strongly held beliefs in the political causes of the FMLN (Viterna 2006: 20). This connection is reflected in the fact that all seven politicized guerrillas described previous involvement in organizations advocating reform (previnv) and explained their recruitment to the FMLN movement through these networks. Viterna also notes that five of the seven politicized guerrillas had family members who were active in the FMLN—another network connection (activefam). However, she describes the biographical details of the seven guerrillas as varied.

Using generalized AI, it is possible to examine combinations of background and network conditions and thereby to identify "modal configurations"—widely shared combinations of conditions. For this analysis two conditions, rcamp and complete, are not used because the data on these two conditions is inconsistent with substantive and theoretical expectations. Refugee camps (rcamp) offer a networking venue, but rcamp = 0 for all politicized guerrillas. Having a complete family (complete) is expected to hinder mobilization, but six out of seven politicized guerrillas have complete = 1.

Table 8-4 presents the conversion of table 8-3 into causal recipes, accomplished in three steps. First, cases are sorted into rows based on their profiles. Second, dichotomized conditions are transformed from "present versus absent" codings to "contributing versus irrelevant." The revised codings are based on substantive and theoretical knowledge. For example, the *absence* of family members active in the FMLN is not considered a condition that contributes to joining the FMLN. Dashes are used in the table to indicate irrelevance (see chapter 6). Third, low-frequency combinations ($N < 2$) are dropped from the table, a step that is motivated by the focus on widely shared combinations of antecedent conditions.

The two recipes shown in table 8-4 can be reduced to one, because the first recipe is a logical subset of the second. Joining these two rows yields a single recipe covering five of the seven cases (71.4 percent). The modal configuration for politicized guerrillas is

previnv•young•early•~mother → politicized

Here and below, an arrow indicates the superset/subset relation, a multiplication sign indicates logical *and* (combined conditions), and a tilde indicates *not* (set negation). In other words, the modal politicized guerrilla was a young woman, not yet a mother, who—based on her prior involvement in oppositional organizations—joined the FMLN during the early years of the struggle.

RELUCTANT GUERRILLAS

Table 8-5 summarizes Viterna's tabular data on the fourteen reluctant guerrillas included in her sample. These women joined and worked in the guerrilla camps because they had no other option. Each woman faced a life-threatening crisis in the early, more violent years of the war and was unable to escape to a refugee camp. Many had family members who were active in the FMLN, which may have facilitated their absorption into the guerrilla camps. The conditions listed in table 8-5 provide several important leads for specifying modal combinations. Most reluctant guerrillas joined the guerrilla camps during the early years of the war (early); most did not have the resources of a complete family (complete); by definition, none of the reluctant guerrillas resided in refugee camps (rcamp); and many had family members active in the FMLN (activefam).

Table 8-6 summarizes the conversion of the tabular data, just described, into recipes. Again, there are three steps: (1) sorting the cases according to their profiles of conditions; (2) converting "presence versus absence" conditions into "contributing versus irrelevant" conditions, based on substantive and theoretical knowledge; and (3) deleting low-frequency recipes ($N < 2$). The two final recipes are listed in table 8-6. The first listed recipe is a logical subset of the second. Thus, the table reduces to a single modal configuration. Note that the fourteen reluctant guerrillas all experienced life-threatening crises, which should be considered part of the modal configuration, even though it is not listed by Viterna as a condition in her tabular data:

early•~rcamp•activefam (•crisis) → reluctant

This combination embraces eleven of the fourteen reluctant guerrillas, a coverage of 78.6 percent. The recipe reflects that these women became guerrillas, reluctantly,

TABLE 8-5 Tabular data on reluctant guerrillas

name	young	early	mother	complete	rcamp	previnv	activefam
Julia	1	1	0	0	0	0	1
Claudia	1	1	0	0	0	0	1
Maria	1	1	0	0	0	0	1
Yenifer	1	1	0	0	0	1	1
Blanca	1	1	0	1	0	0	1
Juana	0	0	1	0	0	1	1
Gladis	0	1	1	0	0	1	1
Lulu	0	1	1	0	0	1	1
Angela	0	1	1	0	0	1	1
Margarita	0	1	1	1	0	1	1
Mirna	0	1	1	0	0	0	1
Rosmaria	0	1	1	0	0	0	1
Yaniris	0	1	1	0	0	0	0
Andrea	0	1	1	1	0	0	0

TABLE 8-6 Data on reluctant guerrillas converted to recipes

early	complete	rcamp	activefam	number
1	0	0	1	9
1	–	0	1	2

in the early, more violent years of the war, were unable to take shelter in the refugee camps, and often relied on family members who were active in the FMLN.

RECRUITED GUERRILLAS

In the later period of the war, FMLN activists visited the refugee camps in a concerted effort to recruit women to become guerrillas. Viterna (2006: 31) notes that the women "were not invited to participate because they shared common ideologies with the guerrillas, but rather were identified by their perceived biographical availability." Recruiters targeted young, childless women, who had "incomplete" families. These women had fewer barriers to participation (2006: 30) and were seen as prime candidates for recruitment. Table 8-7 summarizes Viterna's tabular data on the seventeen recruited guerrillas included in her sample. The commonalities shared by the seventeen recruited guerrillas reflect the fact that the recruiters had a specific profile in mind: most were young, most were recruited during the second phase of the war, most were not mothers, most lived in refugee camps, and most had family members active in the FMLN.

TABLE 8-7 Tabular data on recruited guerrillas

name	young	early	mother	complete	rcamp	previnv	activefam
Marlene	1	0	0	0	1	0	1
Rebecca	1	0	0	1	1	0	1
Elsy	1	0	0	0	1	0	1
Bellini	1	0	0	0	1	0	1
Minta	1	0	0	0	1	1	1
Sury	1	0	0	0	1	1	1
Aracely	1	0	0	0	1	0	0
Candelaria	1	0	0	1	1	0	1
Leonora	1	0	0	0	0	0	1
Marta	1	0	0	0	1	0	1
Lorena	1	0	0	0	1	0	1
Dolores	1	0	0	0	1	0	1
Lupe	1	0	0	0	1	0	1
Amarenta	1	0	0	0	1	0	0
Magaly	1	0	0	0	1	0	1
Yamileth	1	0	0	0	1	0	0
Rosa	0	1	1	1	1	0	1

TABLE 8-8 Data on recruited guerrillas converted to recipes

young	early	mother	complete	rcamp	activefam	number
1	0	0	0	1	1	10
1	0	0	0	1	–	3
1	0	0	–	1	1	2

Table 8-8 summarizes the conversion of the tabular data, just described, into recipes. As with politicized and reluctant guerrillas, there are three steps to the conversion: (1) sorting the cases according to their profiles of conditions; (2) converting "presence versus absence" conditions into "contributing versus irrelevant" conditions, based on substantive and theoretical knowledge; and (3) deleting low-frequency recipes ($N < 2$). The three final recipes are listed in table 8-8. The first listed recipe is a logical subset of the second and also of the third, which reduces the number of recipes to two. The two remaining recipes are almost identical. One is young•~early•~mother•rcamp•~complete; the other is young•~early•~mother•rcamp•activefam. They differ on only one condition each (~complete vs. activefam). These two conditions can be seen as "substitutable" (see chapter 7) because

TABLE 8-9 Tabular data on collaborators

name	mother	complete	rcamp	previnv	activefam
Francesca	1	0	1	0	1
Eva	1	0	1	1	1
Susana	1	0	0	1	1
Tina	1	0	0	0	1
Griselda	1	0	0	0	1
Lisa	1	1	1	1	1
Nina	1	1	0	1	1
Nela	0	1	1	1	1
Celestina	0	1	0	0	1
Marina	0	1	0	0	1
Magdalena	0	1	0	1	1
Deisy	0	0	0	0	0

they both refer to family contexts that support recruitment to the guerrilla cause. Thus, the table can be reduced to a single modal configuration:

young•~early•~mother•rcamp•(~complete + activefam) → recruited

Here and below, the plus sign indicates alternate conditions (logical *or*). This modal configuration embraces fifteen of the seventeen recruited guerrillas, a coverage of 88.2 percent.

COLLABORATORS

Collaborators were women who took on the risk of having a defined support role for the guerrillas, but who maintained a primary place of residence away from the camps. Their shared characteristics were few. Viterna describes two main types of collaborators: (1) mothers with incomplete families (typically, a single mother with young children) and (2) young non-mothers living with complete families (typically, both parents). Table 8-9 presents Viterna's tabular data on collaborators. A few in both groups (mothers versus non-mothers) experienced living in refugee camps (rcamps), and a few in both groups had previous involvement in organizations advocating reform (previnv). The only widely shared condition, however, is having family members active in the FMLN.

Table 8-10 summarizes the conversion of the tabular data, just described, into recipes. As with the other analyses, there are three steps to the conversion, described previously. However, a different cutoff value was used to identify low-frequency rows ($N < 3$), in order to better represent Viterna's account of the

TABLE 8-10 Data on collaborators converted to recipes

mother	complete	activefam	N
1	0	1	5
0	1	1	4

conditions relevant to collaborators. The first recipe is specific to mothers with incomplete families (~complete); the second recipe is specific to young non-mothers residing with complete families. Having FMLN-active family members is central to both groups. The two groups can be joined using logical *or*, yielding

activefam•(mother•~complete + ~mother•complete) → collaborators

Note that the two pairs of conditions joined by logical *or* (plus sign) both involve family situations that posed obstacles to joining the guerrilla camps (single mothers or uncooperative parents). The recipe covers nine of the twelve collaborators, which is a coverage of 75.0 percent.

NONPARTICIPANTS

Nonparticipants constitute a large and heterogeneous subset of Viterna's data. In some respects, the diversity of nonparticipant cases confirms the discussion of "heterogeneous complements" presented in chapter 4. Cases that are united only by their failure to exhibit the focal outcome (or any one of several focal outcomes, as in Viterna's study) are likely to be heterogeneous. While it may be possible to determine the types of things that "happened instead" of the focal outcomes, it is often either difficult to do so or simply not a priority, given the primary goal of explaining cases with the focal outcomes. In her brief discussion of nonparticipants, Viterna (2006: 35) focuses on what nonparticipants generally lacked: "Unlike politicized guerrillas, few nonparticipants took part in previous political organizations. Unlike reluctant guerrillas, most nonparticipants with crises had the necessary resources to reach a refugee camp. . . . Unlike recruited guerrillas, most nonparticipants living in refugee camps had a complete family and did not have a history of political involvement."

Delving into the various fates of the thirty-two nonparticipants would have constituted an entirely different investigation, well beyond the scope of Viterna's study. For example, from time to time, some of the nonparticipants collaborated with the guerrillas, but without taking on a formal support role. Others avoided such risks altogether. Some fled to refugee camps in San Salvador, far from the front line of the civil war, and so on.

Still, it is useful to identify general patterns in the data. Viterna's tabular data on the thirty-two nonparticipants is presented in table 8-11. The most striking finding is the widespread absence of previous involvement with political or religious

TABLE 8-11 Tabular data on nonparticipants

name	mother	complete	rcamp	previnv	activefam
Perona	1	1	1	0	1
Prudencia	1	1	1	0	1
Teresa	1	1	1	0	0
Clara	1	0	1	0	1
Virginia	1	0	1	0	1
Elena	1	0	1	0	1
Ines	1	0	1	0	1
Norma	1	0	0	0	1
Nidia	1	1	0	0	1
Flor	1	1	0	1	0
Erlinda	1	0	0	0	1
Morena	1	1	1	0	1
Olga	1	1	1	0	1
Daniela	1	1	1	0	1
Cornelia	1	1	1	0	1
Gilda	1	0	1	0	1
Isabela	1	0	1	0	1
Dorotea	1	1	0	0	0
Doti	1	0	0	0	0
Lola	1	0	0	0	0
Monica	0	1	1	0	1
Feliciana	0	1	1	0	1
Adela	0	0	1	0	1
Concepcion	0	0	1	0	1
Vicenta	0	0	1	0	1
Orbelina	0	0	1	0	1
Ancelma	0	0	1	0	1
Alejandra	0	0	1	0	1
Dina	0	0	1	0	1
Nicolasa	0	0	0	0	1
Dora	0	1	0	0	0
Gabriela	0	0	0	0	0

TABLE 8-12 Data on nonparticipants converted to recipes

mother	rcamp	previnv	number
1	1	0	13
–	1	0	9
1	–	0	6

reform organizations (previnv). Also noteworthy is the high proportion of mothers, a biographical factor known to pose major obstacles to participation. Furthermore, many nonparticipants fled to the safety of the refugee camps, including camps in the capital city San Salvador.

Table 8-12 summarizes the conversion of Viterna's tabular data into recipes. As described previously, there are three steps to the conversion. The cutoff value for low-frequency rows was $N < 4$. The first recipe is a logical subset of the other two recipes. The second and third recipes can be joined using logical *or*, yielding

$$\sim\text{previnv}\bullet(\text{mother} + \text{rcamp}) \rightarrow \text{nonparticipant}$$

The recipe covers twenty-eight of the thirty-two nonparticipants (87.5 percent). It indicates that a lack of previous involvement with reform organizations is the main driver of nonparticipation, especially when combined with a biographical obstacle (i.e., motherhood) or an overarching concern for safety (residence in a refugee camp).

CONCLUSION

Viterna's study of Salvadoran women offers a useful platform for a demonstration of generalized AI. Viterna posits substantial diversity in the outcomes she studies and does not let the guerrilla/non-guerrilla dichotomy straitjacket her analysis. She delineates sharp distinctions among the guerrillas, ranging from those who joined on the basis of prior political commitments, to those who had little or no choice but to join, to those who were recruited. Among the non-guerrillas, a substantial number of collaborators were committed to supporting the FMNL cause but, for various reasons, did not reside in the guerrilla camps. The nonparticipants were diverse, united primarily by their lack of prior involvement in oppositional organizations.

It is important to note that Viterna separates the five outcomes and analyzes them one at a time, a practice consistent with generalized AI. This approach to the evidence allows for the possibility that different conditions may be linked to different outcomes. Furthermore, in some contexts it is possible for a condition to be a contributing factor, while in other contexts the same factor could be an inhibiting factor. As Viterna (2006: 2) states, "the same causal factor that promotes mobilization in some people may actually inhibit mobilization in others."

TABLE 8-13 Coverage of focal outcomes compared to incidence in other outcomes*

Modal configuration	Prevalence in focal outcome	Prevalence in other outcomes
Politicized	0.714 ($N = 7$)	0.040 ($N = 75$)
Reluctant	0.786 ($N = 14$)	0.220 ($N = 68$)
Recruited	0.882 ($N = 17$)	0.154 ($N = 65$)
Collaborators	0.750 ($N = 12$)	0.314 ($N = 70$)
Nonparticipants	0.875 ($N = 32$)	0.420 ($N = 50$)

*Differences in proportions are statistically significant in each row ($p < 0.05$).

The five applications of generalized AI culminated in five modal configurations, with each one covering more than 70 percent of the cases with the outcome in question:

politicized: previnv•young•early•~mother
reluctant: early•~rcamp•activefam [•crisis]
recruited: young•~early•~mother•rcamp•(~complete + activefam)
collaborators: activefam•(mother•~complete + ~mother•complete)
nonparticipant: ~previnv•(mother + rcamp)

The modal configurations for politicized guerrillas and recruited guerrillas are especially well articulated. Politicized guerrillas conform closely to literature-based expectations regarding women who become guerrillas. The recipe for recruited guerrillas reflects the profile expectations of the FMLN recruiters. Table 8-13 contrasts the prevalence of each modal configuration in its focal outcome with its prevalence in the four other outcomes combined. The differences are all substantial and statistically significant, ranging from a 0.728 gap for recruited guerrillas to a 0.455 gap for nonparticipants. In general, larger gaps indicate better-articulated modal configurations.

9

Applying Generalized AI to Conventional Quantitative Data

One of the strengths of Viterna's study (chapter 8) is her integration of qualitative interview data with the examination of cross-case evidence. The qualitative evidence, usually personal narratives, reinforces and enlivens her analysis of cross-case patterns. Rarely do most social scientists have the opportunity to join and triangulate different types of evidence in a single study. The most common situation is for the researcher to have a set of quantitative data on multiple cases, most often at the individual level, and nothing more. Furthermore, the analyst often does not participate in the collection of the data and thus has little opportunity to enrich the quantitative analysis with qualitative evidence.

The purpose of this chapter is to demonstrate that generalized AI can be usefully applied to conventional quantitative data. Because generalized AI is fundamentally descriptive in nature, it can be used to complement findings derived using conventional quantitative methods. Applying different analytic techniques to the same data does not make the research multimethod; however, using multiple analytic techniques allows the researcher to observe the impact of different underlying assumptions on findings, especially when the techniques make contrasting assumptions regarding the nature of causation.

The demonstration of generalized AI presented in this chapter uses data from the National Longitudinal Survey of Youth (NLSY), 1979 sample. The analysis is restricted to Black females and focuses on membership in the set of respondents in poverty as the outcome. Before presenting the application of generalized AI to the NLSY data, I offer two quantitative analyses. The first applies logistic regression techniques to a binary dependent variable, in poverty versus not in poverty. The second analysis parallels the first, except that the dependent and independent variables are operationalized as *fuzzy sets* (see appendix B). The second quantitative analysis uses ordinary least squares (OLS) regression to build a bridge between the logistic regression analysis and the application of generalized

AI to the fuzzy-set data. As discussed previously, AI is fundamentally set-analytic in nature. To utilize the truth table techniques presented in this work, causal conditions must be operationalized as crisp or fuzzy sets. To ensure comparability of results, I use the fuzzy sets that were prepared for the generalized AI application as my dependent and independent variables in the second quantitative analysis.

LOGISTIC REGRESSION ANALYSIS

The first quantitative analysis regresses "poverty status" on three interval/ratio-scale variables (respondent's parents' income-to-poverty ratio, respondent's years of education, and respondent's Armed Forces Qualifying Test percentile score) and two dichotomous variables (married vs. not married, and having one or more children vs. having no children). Details regarding the measures used in the logistic regression are provided in appendix E.

Poverty status is a dichotomy, with 1 indicating that household income is less than or equal to the "poverty level" for households of that type (determined by the number of adults, the number of children, and so on), and 0 indicating that household income is greater than the poverty level. For example, if the respondent's household income is $14,000 for a family of four (two adults and two children), and the poverty level for households of that type is $15,000, the income-to-poverty ratio is 14,000/15,000 = 0.93, which would translate to a score of 1 on poverty status. An income-to-poverty ratio of 1.0 or less indicates that the respondent is in poverty.

Parents' income-to-poverty ratio is constructed in the same manner, as a ratio of household income to poverty level. However, it is entered into the logistic regression analysis as a ratio-scale independent variable, not as a dichotomy. Respondent's years of education is linked to educational degrees—such that, for example, a score of 12 indicates that the respondent completed high school. The Armed Forces Qualifying Test (AFQT) is mistakenly treated as a generic test of intelligence by some researchers (e.g., Herrstein and Murray in *The Bell Curve*). However, it is best viewed as a test of the respondent's trainability, which is how it is used by the military. Basically, it is a measure of the respondent's degree of retention of school-based learning. Thus, it is indirectly a measure of school performance, as well as a measure of the respondent's degree of acquiescence to authority.

Table 9-1 reports the results of the logistic regression of poverty status on the five independent variables. All five have statistically significant effects on the odds of being in poverty for Black females. Having children more than doubles the odds of poverty (odds ratio = 2.171), while being married dramatically reduces the odds (odds ratio = 0.125). Parents' income-to-poverty ratio, respondent's years of education, and respondent's AFQT percentile score all reduce the odds of poverty. Overall, these results are consistent with those reported in Ragin and Fiss (2017) and, more generally, with findings reported in the research literature on poverty.

TABLE 9-1 Logistic regression analysis of poverty status, Black female sample

	Coefficient (standard error)	Odds ratio
Children (1 = yes)	0.775 *** (0.225)	2.171
Married (1 = yes)	−2.083*** (0.244)	0.125
Parents' income-to-poverty ratio	−0.112* (0.045)	0.894
Respondent's years of education	−0.468*** (0.074)	0.627
AFQT percentile score	−0.020** (0.007)	0.980
Constant	5.703*** (0.906)	299.853

NOTES: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; pseudo-$r^2 = 0.285$; likelihood-ratio $\chi^2 = 274.47$ (df = 5); $N = 775$.

OLS REGRESSION ANALYSIS USING FUZZY SETS

The OLS regression analysis that follows serves as a bridge between the logistic regression analysis, just presented, and the application of generalized AI, still to come. The regression analysis is unconventional in that it uses fuzzy sets in place of the more familiar variables used in the logistic regression analysis. Before presenting the results of the OLS regression, I provide an overview of the construction and calibration of the relevant fuzzy sets.

The dependent variable is degree of membership in the set of households in poverty. This fuzzy set uses the following benchmarks to convert a respondent's income-to-poverty ratio into degree of membership in the set of households in poverty:

Income-to-poverty ratio	Poverty membership score
0 to 1	1 to 0.95
1 to 2	0.95 to 0.5
2 to 3	0.5 to 0.05
3+	0.05 to 0

The use of a ratio of three times the poverty level for full membership in the set of cases not in poverty is a conservative cutoff value, but also one that is anchored in substantive knowledge regarding what it means to be out of poverty. For example, in 1989, the weighted average poverty threshold for a family of two adults and two children was about $12,500 (Social Security Bulletin, Annual Statistical Supplement

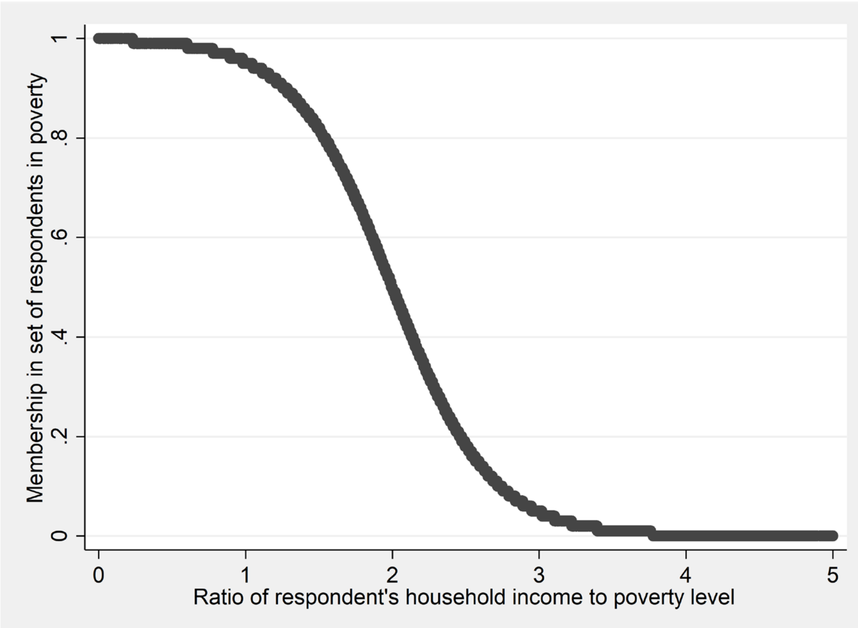


FIGURE 9-1. Calibration of degree of membership in poverty.

1998: tbl. 3.E). Three times this poverty level corresponds to $37,500 for a family of four, a value that lies just slightly above the median family income of $35,353 in 1990 (U.S. Census Bureau, Historical Income Tables—Families, tbl. F-7).

Figure 9-1 illustrates the translation of income ratio values to fuzzy membership scores. For presentation purposes, the *x*-axis has been truncated at an income-to-poverty ratio of 5, consistent with the fact that the threshold for non-membership in the outcome set (an income-to-poverty ratio of 3) has been surpassed by a substantial margin. Note that the calibration of degree of membership in poverty is much more nuanced than the dichotomous dependent variable used in the logistic regression analysis. The dichotomy treats respondents who are barely out of the set of households in poverty (e.g., with an income-to-poverty ratio of 1.01) the same as respondents who are well-off (e.g., with an income-to-poverty ratio of 5 or even greater). The crossover point of the fuzzy set, which separates respondents who are more in versus more out of the set in poverty, is an income-to-poverty ratio of 2.

In place of the two dichotomous variables used in the logistic regression analysis, married versus not married and having children versus not having children, the OLS regression analysis uses a single fuzzy set, *favorable domestic situation*, coded as follows:

Family status combination	Membership in favorable domestic situation
married, no children	1.0
married with children	0.6
unmarried, no children	0.4
unmarried with children	0.0

The membership scores are arrayed according to the association of the categories with poverty. Marriage tends to offer a degree of insulation from poverty, while having children makes poverty more likely. Thus, the highest membership score in the fuzzy set *favorable domestic situation* is for respondents who are married without children (1.0); the lowest membership score is for unmarried respondents with children (0.0). The two middle combinations, *married with children* and *unmarried without children*, both entail domestic situations that are equivocal with respect to poverty avoidance, earning them membership scores close to the crossover point (0.5). However, respondents who are married with children are coded as slightly more in than out of *favorable domestic situation* (0.6), while respondents who are *unmarried without children* are coded as slightly more out than in (0.4).

In place of parents' income-to-poverty ratio, the OLS regression analysis uses *not-low parental income*, a specific calibration of parents' income-to-poverty ratio. The numerator of this measure is based on the average of the reported 1978 and 1979 total net family income in 1990 dollars. The denominator is the household-adjusted poverty level for that household. As explained in appendix B, fuzzy sets use adjectives to specify the range of relevant variation in a source variable. While "parental income" does not make sense as a fuzzy set, "high parental income" and "low parental income" can both be calibrated as fuzzy sets, using data on parental income-to-poverty ratio as the source variable. It is important to note that "low parental income" is not the simple mathematical reverse (i.e., set negation) of "high parental income." A middle-income respondent registers relatively low membership in both "low income" and "high income." The negation of "high income" is "not-high income"; the negation of low income is "not-low income."[1] The benchmarks for degree of membership in *not-low-income parents* are as follows:

Parents' income-to-poverty ratio	Membership in not-low income
0 to 2	0 to 0.05
2 to 3	0.05 to 0.5
3 to 5.5	0.5 to 0.95
5.5+	0.95 to 1

Degree of membership in the set of respondents with *not-low AFQT scores* is based on categories used by the Department of Defense to place enlistees. The military divides the AFQT scale into five categories based on percentiles. Persons in categories I (93rd to 99th percentiles) and II (65th to 92nd percentiles) are considered above average in trainability; those in category III (31st to 64th percentiles) are considered about average; those in category IV (10th to 30th percentiles) are designated as below average in trainability; and those in category V (1st to 9th percentiles) are designated as well below average. To construct the fuzzy set of respondents with *not-low AFQT scores*, I use respondents' AFQT percentile scores. The threshold for full membership (0.95) in the set of respondents with *not-low AFQT scores* was placed at the 30th percentile, in line with its usage by the military; respondents who scored greater than the 30th percentile received fuzzy membership scores greater than 0.95. The crossover point (0.5) was set at the 20th percentile, and the threshold for non-membership was set at the 10th percentile, again reflecting the practical application of AFQT scores by the military. Respondents who scored worse than the 10th percentile received fuzzy scores less than 0.05 in degree of membership in the set of respondents with *not-low AFQT scores*.

AFQT percentile score	Membership in not-low AFQT score
1st to 10th	0 to 0.05
10th to 20th	0.05 to 0.5
20th to 30th	0.5 to 0.95
30th+	0.95 to 1

Respondent's years of education serves as the source variable for the fuzzy set, degree of membership in the set of *educated* respondents. The translation of years of education to fuzzy membership scores is detailed below. Respondents with twelve or more years of schooling are more in than out of the set of educated respondents (fuzzy score > 0.5). Those with fewer than nine years of education are treated as fully out of the set of educated respondents (fuzzy score of 0.0), and those with sixteen or more years of education are treated as fully in the set of educated respondents.

Years of education	Membership in educated
0–8	0.0
9	0.1
10	0.2
11	0.4
12	0.6
13	0.7
14	0.8
15	0.9
16 (max.)	1.0

Table 9-2 reports the results of the OLS regression analysis using fuzzy-set membership scores for the dependent and independent variables. Overall, the results are entirely consistent with the logistic regression analysis reported in table 9-1.

TABLE 9-2 OLS regression analysis of degree of membership in poverty, Black female sample

	Coefficient (standard error)	Standardized coefficient
Favorable domestic situation	−0.487*** (0.036)	−0.366
Not-low parental income	−0.188*** (0.030)	−0.180
Educated	−0.440*** (0.059)	−0.237
Not-low AFQT score	−0.215*** (0.032	−0.216
Constant	1.151*** (0.035)	–

NOTES: *** $p < 0.001$; $r2 = 0.466$; $F = 167.91$ (df = 4 and 770); $N = 775$.

All four independent variables have negative effects on degree of membership in poverty, and all four are statistically significant at $p <$ 001.[2] The metric regression coefficients indicate the decrease in membership in poverty associated with full membership in each of the condition sets. Thus, the four independent variables utilize the same metric. For example, a respondent with full membership in *favorable domestic situation* (i.e., respondent is married and childless) registers a 0.487 decrease in the outcome, degree of membership in poverty. Full membership in the set of *educated* respondents also has a very strong metric effect on membership in poverty, a reduction of 0.440. The r^2 value of this analysis, 0.466, is impressive for individual-level data.[3]

APPLICATION OF GENERALIZED AI TO NLSY DATA

The first step in applying generalized AI to conventional quantitative evidence is to reconceptualize the dependent variable. Instead of being viewed as a raw quantity that simply varies across cases, the dependent variable must be reformulated as one or more qualitative outcomes. Fortunately, this focus on qualitative outcomes is consistent with the logic of the calibration procedure used to create fuzzy sets from conventional ratio- and interval-scale variables. To create a fuzzy set, the researcher specifies numerical values for the two main qualitative breakpoints—the threshold for full membership in the set and the threshold for full non-membership.[4] For example, the calibration of membership in the set of respondents in poverty, described above, uses an income-to-poverty ratio of 1.0 as the threshold for full membership in the set. Respondents with a ratio of 1.0 or less are classified as in poverty. Likewise, the qualitative threshold for non-membership in the set is an income-to-poverty ratio of 3.0. Respondents with a ratio of 3.0 or greater are classified as fully out of poverty. Thus, there is a direct link between generalized AI's focus on qualitative outcomes and the interpretive work that is central to the construction and calibration of fuzzy sets.

From the perspective of generalized AI, there are two key questions addressed by the analysis: (1) What causally relevant conditions are shared by respondents with full membership in the set in poverty? (2) What causally relevant conditions are shared by respondents with full non-membership in this set? Note that these two analyses are independent of each other. In other words, the "negative" cases (i.e., those with non-membership in the outcome set) do not serve as analytic foils for the examination of the positive cases, as they do in the two quantitative analyses presented above. Rather, these "negative" cases are accorded equal analytic attention and are treated as instances of a separate outcome. This feature of generalized AI contrasts sharply with the two quantitative analyses.

The causally relevant conditions under consideration are the four fuzzy sets used in the OLS regression analysis: *favorable domestic situation, not-low-income parents, educated respondent*, and *not-low test score*. Note, however, that it is the absence (or negation) of these conditions that should be linked to membership in the set of respondents in poverty, while their presence should be linked to non-membership in this set. In other words, the interpretive inferences (see chapter 6) that shape the coding of conditions in the two truth tables are opposite.

Table 9-3 presents the results of the application of generalized AI to the set of respondents in poverty (outcome set membership ≥0.95). There are three main steps. First, respondents are sorted into truth table rows based on their profiles. Membership scores greater than 0.5 (the crossover point) are treated as present (1); membership scores less than 0.5 are treated as absent (0). For example, the first row of the table summarizes the eighty-one respondents in poverty who have less than 0.5 membership in three conditions (not-low parental income, not-low AFQT scores, and favorable domestic situation), and greater than 0.5 membership in one—the set of educated respondents. Second, the four conditions are transformed from "present versus absent" codings (panel A) into "contributing versus irrelevant" codings (panel B). The revised codings are based on substantive and theoretical knowledge. For example, the absence of a *favorable domestic situation* is clearly linked to poverty, while its presence is not. Dashes are used in panel B to indicate irrelevance (see chapter 6). Third, low-frequency combinations ($N <$ 20) have been dropped from the table, which is motivated by the focus on the most widely shared combinations of contributing conditions (i.e., "modal configurations"). The three listed rows together embrace 67 percent of the respondents experiencing poverty.

The next step is to simplify the panel B results. In fact, the first and second rows (~nlpinc•~nlafqt•~fdomsit and ~nlpinc•~educ•~nlafqt•~fdomsit) are both

TABLE 9-3 Conditions linked to poverty (frequency cutoff: $N \geq 20$)

Panel A

Not-low parental income (nlpinc)	Educated (educ)	Not-low AFQT (nlafqt)	Favorable domestic situation (fdomsit)	*N*
0	1	0	0	81
0	0	0	0	59
0	1	1	0	23

Panel B

Not-low parental income (nlpinc)	Educated (educ)	Not-low AFQT (nlafqt)	Favorable domestic situation (fdomsit)	*N*
0	–	0	0	81
0	0	0	0	59
0	–	–	0	23

logical subsets of the third row (~nlpinc•~fdomsit). Thus, table 9-3 panel B reduces to a single modal configuration:

~nlpinc•~fdomsit → poverty

Here and below, an arrow indicates the superset/subset relation, a multiplication sign indicates the logical term *and* (combined conditions), and a tilde indicates *not* (set negation). In short, poverty is strongly linked to the combination of low parental income and an unfavorable domestic situation. The other two conditions, being educated and having not-low AFQT scores, are not consistently absent among respondents in poverty. It is important to note, in regard to low parental income and unfavorable domestic situation, that (1) they are conjunctural in the modal configuration, meaning that it is their combination that matters; and (2) both concern family characteristics, in the current household and in the family of origin.

The application of generalized AI to the avoidance of poverty (using the qualitative breakpoint of an income-to-poverty ratio of 3.0 or greater) follows the same general pattern. Table 9-4 panel A shows the high-frequency combinations among the respondents who avoid poverty, along with conventional presence/absence coding of their conditions. Panel B shows the results of the application of interpretive inferences to panel A. Conditions that do not contribute to the outcome are converted into dashes, indicating irrelevance. For example, respondents in the second row of panel A have unfavorable domestic situations, which is not linked to avoiding poverty. Accordingly, this condition is converted into a dash in panel B. Finally, this analysis, like the one preceding it, uses a frequency threshold of twenty respondents, and in so doing embraces 75 percent of the cases avoiding poverty.

TABLE 9-4 Conditions linked to avoiding poverty (frequency cutoff: $N \geq 20$)

Panel A

Not-low parental income (nlpinc)	Educated (educ)	Not-low AFQT (nlafqt)	Favorable domestic situation (fdomsit)	*N*
1	1	1	1	54
1	1	1	0	45
0	1	1	1	33
0	1	1	0	22

Panel B

Not-low parental income (nlpinc)	Educated (educ)	Not-low AFQT (nlafqt)	Favorable domestic situation (fdomsit)	*N*
1	1	1	1	54
1	1	1	–	45
–	1	1	1	33
–	1	1	–	22

Simplifying the results reported in table 9-4 panel B is straightforward. The first three rows are all logical subsets of the fourth row, which leads to a single modal configuration:

$$\text{educ}\bullet\text{nlaftq} \rightarrow \text{avoiding poverty}$$

In other words, avoiding poverty is strongly linked to the combination of being in the set of educated respondents and having not-low AFQT scores. The other two conditions, not-low-income parents and a favorable domestic situation, are not consistently present among respondents avoiding poverty. The results indicate that being educated and retaining school-based learning, the basis for a not-low AFQT score, together offer a degree of protection from poverty, regardless of domestic situation and parental income. The fact that they are conjunctural is consistent with the interpretation that one without the other would not be as effective.

These findings contrast dramatically with the generalized AI results for respondents in poverty. The conditions linked to being in poverty are having an unfavorable domestic situation and low-income parents; low education and low AFQT scores are not consistently linked to poverty. However, as just demonstrated, being educated and not having low AFQT scores are both strongly linked to avoiding poverty. These contrasting findings are not accessible using techniques that merge the two outcomes into a single analysis (i.e., almost all forms of conventional quantitative analysis; see Lieberson 1985). With generalized AI, it is not necessary to use "negative" cases as a foil for the positive cases. The two analyses are separate and equal.

CLARIFYING THE TWO MODAL CONFIGURATIONS

It is important to point out that the two generalized AI solutions, while dramatically different in substance, overlap. This can be verified simply by deriving their intersection. If their intersection produces anything other than a null set, there is logical overlap:

$$(\sim\text{nlpinc}\bullet\sim\text{fdomsit}) \bullet (\text{educ}\bullet\text{nlaftq}) = \sim\text{nlpinc}\bullet\text{educ}\bullet\text{nlaftq}\bullet\sim\text{fdomsit}$$

The overlap occurs in part because the process of applying interpretive inferences eliminates non-contributing conditions on the basis of theoretical and/or substantive knowledge, not on the basis of empirical analysis. Overlap might be acceptable if there were no respondents in the intersection of the two modal configurations (i.e., in the four-way combination just derived). However, as is clear from tables 9-3 and 9-4, there is a nontrivial number of such respondents.

It is a straightforward matter to resolve the overlap, either by awarding it to one of the two modal configurations or by removing it from both, a more conservative strategy. For example, to assign the overlap to the modal configuration for poverty, it is necessary to remove the overlap from the modal configuration for avoiding poverty. The removal can be accomplished by intersecting the modal configuration for avoiding poverty with the *negation* of the modal configuration for poverty. This restricts the modal configuration for avoiding poverty to the combinations of conditions *not covered* by the modal configuration for poverty:

$$\begin{aligned}\text{avoiding poverty} &= (\text{educ}\bullet\text{nlaftq}) \bullet \sim(\sim\text{nlpinc}\bullet\sim\text{fdomsit})\\ &= (\text{educ}\bullet\text{nlaftq}) \bullet (\text{nlpinc} + \text{fdomsit})\\ &= \text{educ}\bullet\text{nlaftq}\bullet\text{nlpinc} + \text{educ}\bullet\text{nlaftq}\bullet\text{fdomsit}\end{aligned}$$

Here and below, a plus sign indicates the logical term *or* (alternate conditions or alternate combinations of conditions). Using De Morgan's theorem, the negation of (~nlpinc•~fdomsit) is (nlpinc + fdomsit). In essence, the scope of the modal configuration for avoiding poverty has been narrowed, while the scope of the modal configuration for poverty (~nlpinc•~fdomsit) is unchanged.

Alternatively, the overlap can be assigned to the modal configuration for avoiding poverty. In this scenario the overlap must be removed from the modal configuration for poverty, which can be accomplished by intersecting it with the negation of the modal configuration for avoiding poverty, as follows:

$$\begin{aligned}\text{in poverty} &= (\sim\text{nlpinc}\bullet\sim\text{fdomsit}) \bullet \sim(\text{educ}\bullet\text{nlaftq})\\ &= (\sim\text{nlpinc}\bullet\sim\text{fdomsit}) \bullet (\sim\text{educ} + \sim\text{nlafqt})\\ &= \sim\text{nlpinc}\bullet\sim\text{fdomsit}\bullet\sim\text{educ} + \sim\text{nlpinc}\bullet\sim\text{fdomsit}\bullet\sim\text{nlaftq}\end{aligned}$$

De Morgan's theorem is applied to educ•nlaftq to produce ~educ + ~nlafqt. The scope of the modal configuration for poverty has been narrowed, while the scope of the modal configuration for avoiding poverty (educ•nlafqt) is unchanged.

Finally, the most conservative strategy is to remove the overlap from both modal configurations, which yields

$$\begin{aligned}\text{in poverty} &= \sim\text{nlpinc}\bullet\sim\text{fdomsit}\bullet\sim\text{educ} + \sim\text{nlpinc}\bullet\sim\text{fdomsit}\bullet\sim\text{nlaftq}\\ \text{avoiding poverty} &= \text{educ}\bullet\text{nlaftq}\bullet\text{nlpinc} + \text{educ}\bullet\text{nlaftq}\bullet\text{fdomsit}\end{aligned}$$

In this version of the results, not being educated or having low AFQT scores accompanies the core conditions linked to poverty (low-income parents combined with an unfavorable domestic situation), and not-low parental income or a favorable domestic situation accompanies the core conditions linked to avoiding poverty (being educated combined with having not-low test scores).

All three solutions to the problem of overlapping solutions are valid. The choice of strategies for addressing the overlap must be based on substantive and theoretical knowledge and interests. For example, if the researcher in this example wanted to emphasize the challenge of avoiding poverty for Black females, she might favor the more restrictive modal configuration for that outcome, and leave intact the less restrictive, two-condition configuration for being in poverty.

DISCUSSION

The findings of the application of generalized AI to the NLSY data on Black females add depth to the results of the two regression analyses. In both regression analyses, independent variables are evaluated with respect to their separate contributions to the explanation of variation in the dependent variable. Variation in the dependent variable is key; without variation, there is nothing to explain. Both analyses confirm that the independent variables all have significant net effects on their respective dependent variables. The application of generalized AI, by contrast, separates the dependent variable into two qualitative outcomes and two separate analyses—full membership in the set of respondents in poverty and full membership in the set of respondents avoiding poverty. The conditions strongly linked to these two outcomes differ: having low-income parents combined with an unfavorable domestic situation is linked to being in poverty; being educated combined with having not-low test scores is linked to avoiding poverty. These are not simple net effects; both solutions involve combinations of conditions.

The two regression analyses and the generalized AI analysis provide convergent results. However, greater nuance is offered by the generalized AI application.[5] So-called independent variables with generic net effects are recast as modal configurations that differ by outcome. The application of generalized AI reveals subtle differences among the four causal conditions. The two conditions that are consistently linked to poverty are inconsistently linked to avoiding poverty, and vice versa. These contrasting effects are masked in the regression analyses.

10

Core Features of Generalized Analytic Induction

This chapter summarizes core features of generalized AI and concludes with a discussion of potential applications. The core features discussed range from its general orientation as a research strategy to practical procedures involved in applying the method. Considered together, these features define a strategy of social inquiry that differs fundamentally from that of conventional quantitative research.

- AI is applied to outcomes that are more or less the same across a range of cases. AI focuses analytic attention on one outcome at a time, and avoids pooling different outcomes in a single analysis. Rather than analyzing a dichotomized outcome as present versus absent, AI emphasizes the separate treatment of each outcome—the focal outcome and substantively important alternate outcomes.
- AI prioritizes the identification of antecedent conditions shared by instances of an outcome. Shared antecedent conditions, in turn, provide a basis for the specification of causal recipes, which in turn serve as guides to causal interpretation at the case level. AI is not an inferential technique; rather, it is largely descriptive and interpretive.
- AI eschews the concept of negative cases, especially when the set of negative cases is defined simply by their failure to display the focal outcome. Negative cases are more appropriately viewed as positive instances of alternate outcomes.
- AI is especially well suited for research questions addressing qualitative outcomes. The guiding question in most such applications of AI is "How did the outcome happen?" "How" questions prioritize positive cases and focus the investigation on combinations of shared antecedent conditions (i.e., "modal configurations").

- AI is dynamic and iterative. The conceptualization of the outcome is open to revision as the investigation proceeds, and the specification of antecedent conditions may be revised as case knowledge deepens. The research process is iterative, as positive/disconfirming cases motivate revisions to the conceptualization of the outcome or to the specification of the working hypothesis.
- AI, especially generalized AI, evaluates the consistency of the set-analytic connection between antecedent conditions and outcomes using enumerative criteria. Generalized AI assesses the degree to which the "inclusion" relation between antecedent conditions and an outcome is satisfied. Classic AI seeks perfect inclusion, with no positive/disconfirming cases remaining at the conclusion of the investigation.
- AI relies heavily on interpretive inferences when assessing antecedent conditions. An interpretive inference recasts a presence-versus-absence dichotomy as a contributing-versus-irrelevant dichotomy, which in turn simplifies the assessment of antecedent conditions. As shown in the applications presented in chapters 6–9, AI's interpretive logic mimics the case-oriented researcher's goal of developing case narratives based on contributing conditions.
- AI's truth table solutions are normally presented in "sum-of-products" form. Converting them into "product-of-sums" form, as demonstrated in chapter 7, can uncover conditions that constitute substitutable ways of satisfying a more general causal requirement. Identifying substitutable conditions can greatly simplify a causal formula. Appendix D describes a procedure for converting a sum-of-products expression into a product-of-sums expression.
- The interpretation of a truth table solution with two (or more) causal recipes can be enhanced by "clarifying" the recipes—assigning the overlap exclusively to one of the recipes and removing it from the other(s). The first step is to determine which recipe is to be awarded the overlap. The second step is to derive the complement (negation) of the selected recipe using De Morgan's theorem. Third, the negation of the recipe is intersected with the alternate recipe, which narrows the breadth of the second recipe while awarding the overlap to the first:

A•B + C•D	two overlapping recipes (overlap: A•B•C•D)
A•B	recipe selected to receive overlap
~(A•B) = ~A + ~B	the complement of the selected recipe
(~A + ~B) • C•D	complement intersected with the second recipe
A•B + C•D•(~A + ~B)	clarified recipes
A•B + C•D•~A + C•D•~B	clarified recipes in sum-of-products form

- When antecedent conditions vary by level or degree, they can be calibrated as fuzzy sets. Once converted into fuzzy sets, they can be utilized as antecedent conditions in truth tables, which sort cases according to their combinations of conditions. The calibration of an interval or ratio-scale variable as a fuzzy set must be grounded in theoretical and substantive knowledge, especially with respect to the crossover point separating cases that are more "in" versus more "out of" the target set (appendix B; see Ragin 2008: chaps. 4 and 5).

POTENTIAL APPLICATIONS

Generalized AI is a flexible tool with many potential applications. This book emphasizes its relevance to "how" questions, where the goal is to identify the antecedent conditions shared by a set of cases with the same outcome. However, generalized AI can be used to address any research question regarding the decisive features or elements shared by the members of a category or set. Consider, for example, the wide array of outcomes, both hypothetical and empirical, mentioned in this work:

becoming a marijuana user
succumbing to opiate addiction
resorting to embezzlement
the rise of modern tyrants
successful local management of common pool resources
the emergence of bureaucratic authoritarian states
the breakdown of democratic regimes
the successful shaming of violators of international agreements
long-term commitment to being an Olympic-caliber athlete
movement organizations that secured advantages for their constituents
being "in" versus "well-out-of" poverty
participation of women in El Salvador's guerrilla army
protesting IMF-mandated austerity measures
engaging in electoral fraud

These outcomes vary on a number of important dimensions. For example, they range in scale from outcomes specific to individuals to outcomes relevant to countries. They also vary in terms of the degree to which they invoke immediate, proximate conditions versus conditions that are more long-term, structural, or contextual in nature. Finally, they vary in terms of their potential for offering findings or conclusions that are transferable to other settings. Some are strongly anchored in specific times and places, while others have wide implications.

As demonstrated in chapter 9, generalized AI can be used in conjunction with variable-oriented methods. Most conventional variable-oriented methods focus on the separate impact of "independent" variables on outcomes. The usual goal is either to gauge the relative importance of competing variables or to demonstrate that a theoretically decisive variable has an independent impact. In either case, the key task is to isolate each independent variable's separate contribution to the outcome. Generalized AI, by contrast, focuses on combinations of contributing conditions—modal configurations. This feature counterbalances the emphasis of the variable-oriented approach on assessing each condition's net contribution to an outcome. Furthermore, by highlighting combinations of conditions, generalized AI provides a bridge to causal interpretation. Combinations of conditions are often suggestive of causal mechanisms, which, in turn, can be explored and assessed at the case level (Goertz 2017).

Generalized AI also can be used in conjunction with case-oriented methods, especially those that examine multiple instances of a qualitative outcome. Many applications of case-oriented methods culminate in a "composite portrait" of such instances. The researcher constructs a representation of the category based on common features. For example, a researcher might construct a composite portrait of committed environmental activists based on interviews with a sample of activists. Very often, the composite portrait that results is an amalgamation of noteworthy features of selected instances, chosen because of their salience to the researcher. Generalized AI makes the process of constructing representations of cross-case evidence both systematic and transparent. By applying the same analytic frame to each case (via truth tables) and directly assessing the degree to which combinations of features are shared across cases, generalized AI brings rigor to a research approach that is often seen as ad hoc.

Generalized AI also aids process tracing, an important case-oriented research tool. One of the central goals of process tracing is to gather case-level evidence relevant to causal mechanisms (Goertz 2017; Schneider and Rohlfing 2016). Often, researchers posit mechanisms based on cross-case analysis and then process trace at the case level as a way to assess the inferred mechanism (Goertz and Haggard 2022). As noted previously, generalized AI focuses on *combinations* of causally relevant antecedent conditions, which in turn are suggestive of causal mechanisms. In addition to offering greater guidance to the effort to identify mechanisms, generalized AI also can be used to aid the selection of cases for in-depth, process-oriented examination. Thus, generalized AI formalizes and systematizes basic analytic strategies commonly practiced by qualitative researchers.

APPENDIX A

Brief Overview of Qualitative Comparative Analysis

Qualitative comparative analysis (QCA) is designed for the investigation of configurations of causally relevant conditions across positive and negative instances of an outcome. An especially useful feature of QCA is its capacity for analyzing complex causation, defined as a situation where an outcome may follow from several different combinations of causal conditions—that is, from different causal "recipes." For example, a researcher may have good reason to suspect that there are several different causal recipes for the consolidation of third wave democracies. By examining the fate of cases with different configurations of causally relevant conditions, across both successful and unsuccessful cases of consolidation, it is possible to identify the decisive recipes and thereby unravel causal complexity.

The key analytic tool for analyzing causal complexity is the truth table, a tool that allows "structured, focused comparisons" (George 1979). Truth tables list the logically possible combinations of causal conditions and the empirical outcome associated with each combination. For example, based on theoretical and substantive knowledge, a scholar might argue that a key recipe for democratic consolidation involves the following combination of conditions: a presidential form of government, a strong executive, a low level of party fractionalization, and a noncommunist past. Table A-1 illustrates a hypothetical truth table operationalizing this argument. With four causal conditions there are sixteen logically possible combinations of conditions (causal configurations). In more complex analyses, the rows (representing combinations of causal conditions) may be quite numerous because the number of causal combinations is a geometric function of the number of causal conditions (number of causal combinations = 2^k, where k is the number of causal conditions). It is important to point out that the procedures described here are not dependent on the use of dichotomies. Truth tables can be built from fuzzy sets (where memberships in sets range from 0 to 1) without dichotomizing the fuzzy scores. These procedures take full advantage of the graded membership scores central to the fuzzy-set approach (see Ragin 2000, 2008; Ragin and Fiss 2017).

TABLE A-1 Hypothetical truth table showing causal conditions relevant to democratic consolidation

Presidential form	Strong executive	Low party fractionalization	Noncommunist	Consolidated
No	No	No	No	–
No	No	No	Yes	No
No	No	Yes	No	No
No	No	Yes	Yes	–
No	Yes	No	No	No
No	Yes	No	Yes	No
No	Yes	Yes	No	–
No	Yes	Yes	Yes	Yes
Yes	No	No	No	No
Yes	No	No	Yes	–
Yes	No	Yes	No	–
Yes	No	Yes	Yes	–
Yes	Yes	No	No	Yes
Yes	Yes	No	Yes	Yes
Yes	Yes	Yes	No	–
Yes	Yes	Yes	Yes	Yes

The use of truth tables to unravel causal complexity is described in detail elsewhere (e.g., Ragin 1987, 2000, 2008; Schneider and Wagemann 2012). The essential point is that the truth table elaborates and formalizes one of the key analytic strategies of comparative research: examining cases that share specific combinations of causal conditions to see if they share the presence or the absence of the outcome. Indeed, the main goal of truth table analysis is to identify connections between combinations of causal conditions and outcomes. By listing the different logically possible combinations of conditions, it is possible to assess not only the sufficiency of a specific recipe (e.g., the recipe presented in the last row of table A-1, with all four causal conditions present), but also the sufficiency of the other logically possible combinations of conditions that can be constructed from these causal conditions. For example, if the cases with all four conditions present experience democratic consolidation and the cases with three of the four conditions present (and one absent) also experience consolidation, then the researcher may conclude that the causal condition that varies across these two combinations is irrelevant to the recipe. The key ingredients for the outcome are the remaining three conditions. Various techniques and procedures for logically simplifying patterns in truth tables, in addition to the simple one just described, are detailed in Ragin (1987, 2000, 2008).

Often, the move from a hypothesized recipe to a truth table stimulates a reformulation or expansion of a recipe, based on an examination of relevant cases. For example, suppose that the truth table revealed substantial inconsistency in the last row—that is, suppose there were a few cases in the last row that failed to consolidate, in addition to the several that did consolidate. This inconsistency in outcomes signals to the investigator that more in-depth study of cases is needed. For example, by comparing the cases in this row that failed to consolidate with those that consolidated, it would be possible to elaborate the recipe. Suppose this comparison revealed that the cases that failed to consolidate all had severe elite divisions. This ingredient (elite divisions) could then be added to the recipe, and the truth table could then be respecified with five causal conditions (and thus thirty-two rows).

The task of truth table refinement is demanding, for it requires knowledge of cases and many iterations between theory, cases, and truth table construction. In effect, the truth table disciplines the research process, providing a framework for comparing cases as configurations of similarities and differences while exploring patterns of consistency and inconsistency with respect to case outcomes.

APPENDIX B

Fuzzy Sets

Degree of membership in a fuzzy set ranges from 0 to 1, with 0 indicating that a case is completely out of the set in question, and a score of 1 indicating full membership in the set. A value of 0.5 is the crossover point, which indicates maximum ambiguity in whether a case is more in or more out of the set in question. As explained in Ragin (2008) and Ragin and Fiss (2017), fuzzy-set membership scores are interpretive in nature. Full membership, full non-membership, and the crossover point are qualitatively derived empirical anchors, based on substantive and/or theoretical knowledge and not on statistical properties of the data (e.g., the mean and standard deviation of the source variable).

Fuzzy sets can be created from interval- and ratio-scale source variables using a procedure called calibration. The researcher specifies three breakpoints in the range of the source variable: the threshold for full membership, the crossover point, and the threshold for full non-membership in the target set. Values of the source variable that are greater than the threshold for full membership are arrayed between 0.95 and 1.0; values of the source variable that are less than the threshold for full non-membership are arrayed between 0.05 and 0. The values between the two thresholds are arrayed between 0.05 and 0.95, with the designated crossover point fixed at a fuzzy membership score of 0.5 (see Ragin 2008: chaps. 4 and 5).

The conceptualization and labeling of the target fuzzy set is of utmost importance. Consider the fuzzy set of high-income respondents versus the fuzzy set of not-low-income respondents. Both fuzzy sets would use respondent's income as the source variable, and both sets would array membership scores according to income level. However, the threshold for full membership in not-low-income would be much lower than the threshold for full membership in high-income (e.g., $45,000 vs. $100,000). Thus, the adjectives attached to fuzzy sets have a pivotal impact on the specification of the three empirical anchors used to calibrate set membership. In effect, these adjectives highlight different ranges of variation in the source variable. For example, variation in income that is well above the threshold for full membership could be defined as irrelevant and therefore truncated to full membership (1). The dependence of the calibration procedure on meaningful empirical anchors provides an interpretive foundation to the operationalization and use of fuzzy sets.

The utility of fuzzy sets is twofold. First is the simple fact that fuzzy sets allow specification of the degree of membership of cases in sets. Conventional binary sets permit only two values, 1 (membership) and 0 (non-membership). Fuzzy sets, by contrast, permit as much fine-grained differentiation of cases as is provided by interval- and ratio-scale variables. The second is that almost all set operations that are associated with conventional binary sets (e.g., intersection, union, negation, subsets, and supersets) can be performed using fuzzy sets. For example, if fuzzy set X is a consistent superset of fuzzy set Y, then X may be interpreted as a shared antecedent condition for Y, assuming that the interpretation is supported by theoretical and substantive knowledge.

APPENDIX C

Using fsQCA Software to Implement Generalized AI

Generalized AI, like fuzzy-set qualitative comparative analysis (fsQCA), is set analytic in nature. Consequently, the two approaches share many operations and procedures. For example, both techniques utilize truth tables to simplify and model combinations of conditions, and both can work with both crisp and fuzzy sets. Thus, it is appropriate (and expedient!) to implement generalized AI as part of the fsQCA package. Generalized AI is implemented in fsQCA beginning with version 4.0, which can be freely downloaded from www.fsqca.com.

The purpose of this appendix is to provide practical instructions regarding the application of generalized AI. For this demonstration, I use the data on social movement organizations ("challengers") published by William Gamson (1990) in *The Strategy of Social Protest*. Gamson developed a sampling frame for social movement organizations (SMOs) in the United States from 1800 to 1945. Table C-1 presents Gamson's raw data for the twenty-six SMOs that gained new advantages for their constituents within fifteen years of their period of activism. The presence/absence conditions and the outcome are coded in Gamson (1990: 36) as follows:

burorgiz: 1 = the challenging group developed a bureaucratic organizational structure; 0 = the group lacked a bureaucratic structure

lowstatus: 1 = the challenging group's constituency was low status (e.g., workers, minorities); 0 = constituency was not low status

displace: 1 = the challenging group's goal was to displace a person in a position of power; 0 = non-displacement goals

help: 1 = the challenging group received help from an outsider (e.g., from another challenging group); 0 = no help from outsiders

acceptnc: 1 = the challenging group won general acceptance as a representative of its constituents; 0 = did not win acceptance

newadv: 1 = new advantages accrued to the challenger's constituency within fifteen years of the challenger's activism; 0 = no new advantages

TABLE C-1 Raw data on challengers that secured new advantages

burorgiz	lowstatus	displace	help	acceptnc	newadv
0	0	0	0	0	1
0	0	0	0	0	1
0	0	0	0	1	1
0	0	0	0	1	1
0	0	0	1	0	1
0	0	0	1	0	1
0	0	0	1	1	1
0	0	0	1	1	1
0	0	1	0	1	1
0	1	0	1	1	1
0	1	0	1	1	1
1	0	0	0	1	1
1	0	0	1	0	1
1	0	0	1	1	1
1	1	0	0	0	1
1	1	0	0	1	1
1	1	0	0	1	1
1	1	0	0	1	1
1	1	0	0	1	1
1	1	0	0	1	1
1	1	0	0	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1
1	1	0	1	1	1

The outcome, new advantages, is constant across the twenty-six cases (all are coded "1" on the outcome), consistent with generalized AI's focus on investigating one well-specified outcome at a time.[1]

The first step is to make sure the data set is in a proper format for the software. In general, it is best to use simple variable names (three to ten characters), avoiding punctuation, dashes, underscores, and embedded spaces. Data should be numeric, with the exception of a column of case names (often the first column). While data can be entered directly into fsQCA, it is usually easier to use Microsoft Excel for data entry, saving the file in *.csv format. Sometimes Excel attaches a blank line at the bottom of the data file. This line must be deleted once the *.csv file is opened in fsQCA. Save the data file after deleting the blank line—if Excel has inserted one. Move the cursor to the blank line; click "Cases" and then "Delete."

The software has two main windows, which are opened at startup. The left window displays the data spreadsheet; the right window displays results. To retrieve a data file, click "File" then "Open" (to input data directly into the program, consult the fsQCA manual—downloadable from www.fsqca.com). In addition to *.csv files, fsQCA also can read tab delimited files (*.dat) and space delimited files (*.txt). Sometimes it is necessary to change the three-letter filename extension to make the data file recognizable by fsQCA.

The software offers a variety of data and case functions for manipulating the contents of the data file. For example, there is a calibration procedure for converting interval- and ratio-scale variables into fuzzy sets. This function is very useful when working with conventional survey or archival data (see, e.g., chapter 9). In order to utilize the truth table function, which is central to both generalized AI and QCA, it is necessary for the causally relevant conditions to be crisp sets (i.e., conventional binary variables) or fuzzy sets. The two types are often mixed in the same analysis. The demonstration presented in this appendix uses all crisp sets.

After retrieving the data file, open the generalized AI dialogue box by clicking "Analyze," and then "Analytic Induction." Figure C-1 shows a screenshot of the initial generalized AI dialogue box, with all the variables listed on the left. The first task is to define the outcome, which, in this example, is new advantages (newadv). Click the outcome variable and then click "Set." Notice that there are several ways (=, <, >, ≤, and ≥) to code the outcome, which becomes a constant value of 1 across the selected cases. In this example, the outcome coding is simply newadv = 1, as shown in figure C-2. It is possible, however, to use interval- and ratio-scale variables to define qualitative outcomes. For example, in chapter 9 the first generalized AI application used an income-to-poverty ratio ≤1.0 to select the relevant cases and to code them with a constant value of 1 on the outcome.

Selecting and coding the antecedent conditions comes next. Click relevant conditions one at a time, followed by "Add." Each condition comes with click boxes for "present" versus "absent," the purpose of which is to implement the interpretive coding of causal conditions, as described in chapter 6 and utilized in chapters 7–9. The user clicks "present" if she expects the condition to contribute to the outcome when the condition is present, and "absent" if she expects the condition to contribute to the outcome when the condition is absent. If neither option is selected, the interpretation is that the condition could contribute when it is either present or absent, depending on context (i.e., what other conditions are present).

Figure C-2 shows the interpretive coding of the five antecedent conditions. Bureaucratic organization (burorgiz) is coded present, based on the literature on social movement organizations. Low-status constituents (lowstatus) remains unspecified, consistent with an expectation that its role as a contributing condition is dependent on context. This way of coding lowstatus allows for the possibility that the contributions of the other antecedent conditions may differ depending on whether the challenger's constituency is low status.[2] Having displacement as a primary goal (displace) is coded as contributing when absent for the simple reason that it is very rarely successfully achieved. Receiving help from outsiders

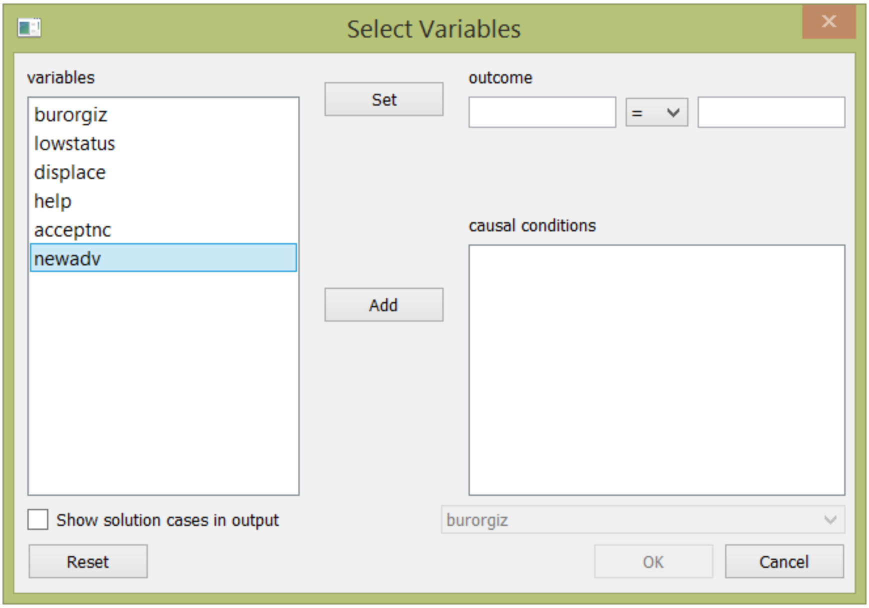


FIGURE C-1. Initial AI dialogue box.

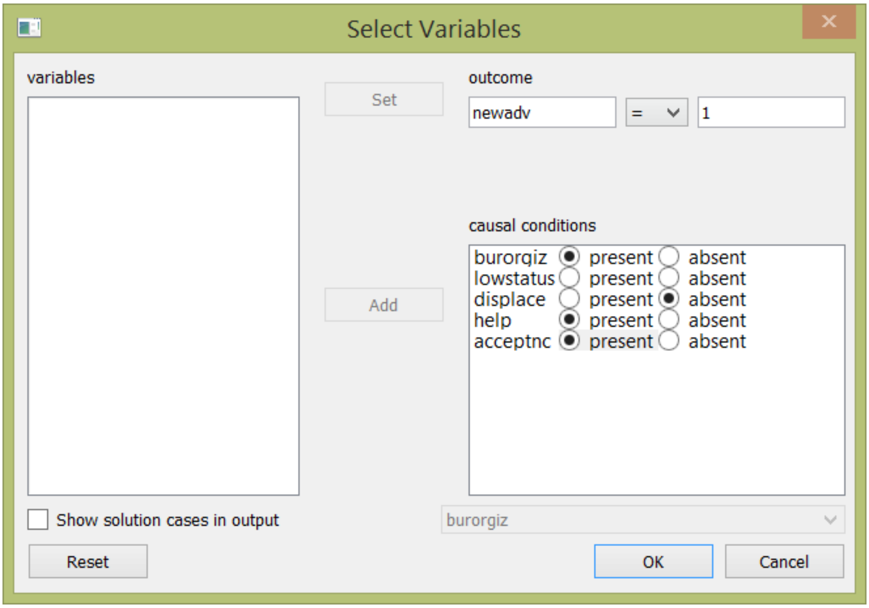
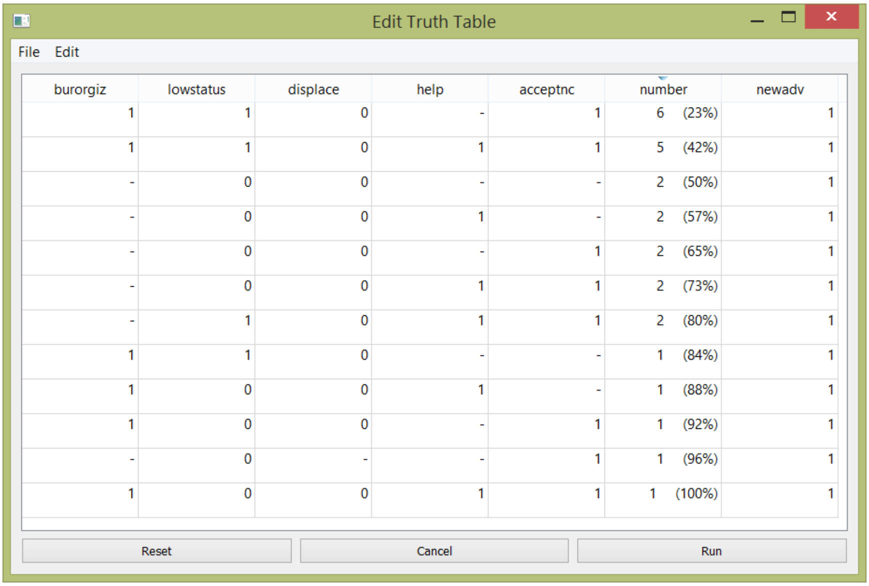


FIGURE C-2. Coded AI dialogue box.



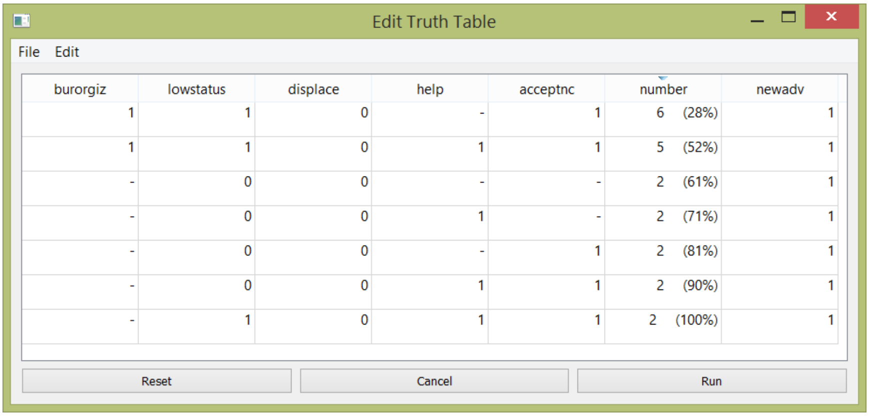
burorgiz	lowstatus	displace	help	acceptnc	number	newadv
1	1	0	-	1	6 (23%)	1
1	1	0	1	1	5 (42%)	1
-	0	0	-	-	2 (50%)	1
-	0	0	1	-	2 (57%)	1
-	0	0	-	1	2 (65%)	1
-	0	0	1	1	2 (73%)	1
-	1	0	1	1	2 (80%)	1
1	1	0	-	-	1 (84%)	1
1	0	0	1	-	1 (88%)	1
1	0	0	-	1	1 (92%)	1
-	0	-	-	1	1 (96%)	1
1	0	0	1	1	1 (100%)	1

FIGURE C-3. Initial truth table.

(help) and achieving acceptance (acceptnc) as a representative of its constituents are both coded as contributing when present, as indicated in the research literature on social movement organizations.

The setup for the analysis is complete. To produce the truth table based on the setup, click “OK” and the truth table window opens, as shown in figure C-3. Cases are sorted into rows based on their condition profiles, and rows are listed in order of the number of cases in each row, as shown in the “number” column. For example, there are six instances of the first row, cases that combine bureaucratic organization, low-status constituency, non-displacement goals, and acceptance. The percentages in the “number” column refer to the cumulative percentage of cases. The dashes in the table indicate that a condition does not contribute to the outcome, based on the interpretive codings input by the user. As explained in chapter 6, “contributing versus irrelevant” codings are based on substantive and theoretical knowledge. Note that the outcome is coded 1 for every row, consistent with generalized AI’s focus on cases sharing a specific outcome.

The next step is to select a meaningful frequency threshold, which determines which rows are included in the logical minimization of the truth table. Generally, the threshold should not be so high that many cases are excluded from the logical minimization. The threshold also should not be too low, which might give too much analytic weight to rows that are deviant in some way or perhaps that exist simply due to classification or measurement error.[3] In this example, I use a frequency threshold of at least two cases, which embraces 80 percent of the cases in the truth table. To implement a numerical threshold, click “Edit” and then “Delete.” A small dialogue box will open, with the message “Delete rows with number less than ____”. In this example, the input value is 2. Then click “OK.” The resulting truth table is shown in figure C-4.



burorgiz	lowstatus	displace	help	acceptnc	number	newadv
1	1	0	-	1	6 (28%)	1
1	1	0	1	1	5 (52%)	1
-	0	0	-	-	2 (61%)	1
-	0	0	1	-	2 (71%)	1
-	0	0	-	1	2 (81%)	1
-	0	0	1	1	2 (90%)	1
-	1	0	1	1	2 (100%)	1

FIGURE C-4. Edited truth table.

TABLE C-2 Causal recipes for new advantages

Causal recipe	Raw coverage	Unique coverage
~lowstatus*~displace	0.423	0.269
burorgiz*~displace*acceptnc	0.5	0.231
~displace*help*acceptnc	0.385	0.077
solution coverage 0.923		

The truth table is now ready for logical minimization. Click "Run." The results are displayed in the output window, which was opened at startup. Table C-2 shows the results of the application of the truth table algorithm. In this example, there are three modal configurations linked to new advantages: (1) challengers with not-low-status constituents combined with non-displacement goals, (2) bureaucratically organized challengers with non-displacement goals and acceptance, and (3) challengers that have achieved acceptance combined with non-displacement goals and help from outsiders. The first modal configuration is found in eleven of the twenty-six instances of new advantages (42.3 percent); its unique coverage (i.e., not overlapping with the coverage of the other two modal configurations) is 26.9 percent (seven of twenty-six instances). The second modal configuration is found in thirteen of twenty-six cases (50 percent); its non-overlapping coverage is six instances (23.1 percent). Finally, the third modal configuration is found in ten cases, with two cases non-overlapping. Altogether, the three modal configurations account for 92.3 percent of the instances of new advantages (twenty-four of twenty-six cases).

CLARIFYING THE TRUTH TABLE RESULTS

While it is tempting to view the results of the truth table algorithm as the conclusion of the analysis, it is important to interrogate the results further. After all, the results can be expressed as a Boolean equation, which in turn can be manipulated algebraically.

First, consider the results expressed as a logical equation:

~lowstatus•~displace + burorgiz•~displace•acceptnc +
~displace•help•acceptnc → newadv

The arrow indicates the superset/subset relation, the multiplication sign indicates the logical term *and* (combined conditions), the plus sign indicates the logical term *or* (alternate conditions or alternate combinations of conditions), and the tilde indicates *not* (set negation). Note that the second and third modal configurations apply to challengers representing low-status constituencies (lowstatus) and also to challengers representing constituencies that are not low status (~lowstatus). Thus, these two recipes and the equation for new advantages can be rewritten as follows:

~lowstatus•~displace + lowstatus•burorgiz•~displace•acceptnc +
~lowstatus•burorgiz•~displace•acceptnc + lowstatus•~displace•help•acceptnc +
~lowstatus•~displace•help•acceptnc → newadv

The next step is important. Two of the terms just added are subsets of the first modal configuration. Specifically, ~lowstatus•burorgiz•~displace•acceptnc is included in ~lowstatus•~displace, and ~lowstatus•~displace•help•acceptnc is also included in ~lowstatus•~displace. Removing the redundant terms yields

~lowstatus•~displace + lowstatus•burorgiz•~displace•acceptnc +
lowstatus•~displace•help•acceptnc → newadv

And then, joining the second and third modal configurations yields

~lowstatus•~displace +
lowstatus•~displace•acceptnc•(help + burorgiz) → newadv

The clarified results reveal that there is an important difference between challengers representing low-status constituents and challengers representing constituents who are not low status. If the constituents are not low status, the only ingredient needed for success is a non-displacement goal. However, if the challenger's constituency is low status, then not only must challengers avoid displacement goals, but they must also win acceptance and either have a bureaucratic organization or help from outsiders. In short, the path to new advantages is much narrower for challengers representing low-status constituents.

Note also that by clarifying the results in this manner, there is no longer overlapping coverage. The first modal configuration covers eleven cases; the second covers thirteen. Total coverage is the same as before: 24/26 (92.3 percent).

APPENDIX D

Converting "Sum-of-Products" Expressions to "Product-of-Sums" Expressions

This appendix presents detailed instructions for converting a sum-of-products expression into a product-of-sums expression, using as an example the analysis presented in chapter 7. Often, the product-of-sums expression will provide important clues regarding which conditions may be substitutable. When two (or more) conditions are substitutable, they can be joined by logical *or* to create a more encompassing condition that is more consistently linked to the outcome than either of the component conditions considered separately. Also, when two conditions are joined by logical *or*, their union typically entails movement to a more abstractly formulated antecedent condition (see chapter 7, and the discussion surrounding table 2-2).

Using fsQCA, there are seven steps to converting a sum-of-products expression into a logically equivalent product-of-sums expression.

1. Derive the truth table solution in sums-of-products form (which is the default form). In chapter 7, the solution is

exercise•feel•rituals + exercise•rituals•assoc +
exercise•feel•food + exercise•assoc•food

(Multiplication indicates combined conditions—set intersection; addition indicates alternate combinations of conditions or alternate conditions—set union.)

2. Click "File," then "New From Expression." An input box will open. Enter the solution from step 1 followed by pressing "Enter" (note the use of asterisks to denote intersection):

exercise*feel*rituals + exercise*rituals*assoc +
exercise*feel*food + exercise*assoc*food

3. Click "Analyze," then click "Truth Table Algorithm." A dialogue box will open. Click "y" (the outcome), then click "Set Negated." Click the five causal conditions one at a time, followed by "Add." Click "OK." The Truth Table window will open.
4. Click "Edit," then "Delete and code." Accept the default settings by clicking "OK."
5. Click "Standard Analyses." The Intermediate Solution dialogue box will open. Click "OK."
6. Switch to the output window (next to the data spreadsheet). The solution reported is

~exercise + ~rituals*~food + ~feel*~assoc
(note the use of the tilde to indicate negation)

7. Derive the complement of the solution by applying De Morgan's theorem to the truth table solution in step 6: intersection is recoded to union, and vice versa. Presence is recoded to absence, and vice versa. The complement generated by applying De Morgan's theorem is the solution in product-of-sums form, logically equivalent to the sum-of-products form, shown in step 1.

Solution (step 6): ~exercise + ~rituals*~food + ~feel*~assoc
Complement: exercise • (rituals + food) • (feel + assoc)

The product-of-sums form—which, in this example, is much simpler than the sum-of-products form—reveals the substitutability of "rituals" and "food," and of "feel" and "assoc."

APPENDIX E

Measures Used in Logistic Regression Analysis

POVERTY STATUS

NLSY data on the official poverty threshold include two measures—poverty level and poverty status—both of which are based on official poverty thresholds (see NLSY79 User's Guide 1999: 240–41). *Poverty level* is the level of income below which a family is considered to be in poverty, adjusted for family size, family composition, and state of residence. It is based on the yearly poverty income guidelines issued by the U.S. Department of Human Services and on Census Bureau poverty guidelines. *Poverty status* is a binary variable that gives the actual status of a family—whether family income is below the poverty threshold—and is calculated from information on poverty level and total family income for the past calendar year. Most researchers use poverty status as a binary dependent variable in logistic regression analyses.

PARENTS' INCOME-TO-POVERTY RATIO

Parental income is assessed by computing the ratio of parental income to the household-adjusted poverty level for the parents' household. The numerator of this measure is based on the average of the reported 1978 and 1979 total net family income in 1990 dollars. The denominator is the household-adjusted poverty level for that household.

RESPONDENT'S YEARS OF EDUCATION

The logistic regression analysis of the number of years of education is adjusted so that twelve years indicates completion of high school, sixteen years indicates completion of a bachelor's degree, and so on.

AFQT PERCENTILE SCORE

AFQT percentile scores are based on the Armed Services Vocational Aptitude Battery (ASVAB), which was introduced by the Department of Defense in 1976 to determine eligibility for enlistment, an issue that had become prominent because of concerns over the quality of recruits as the United States moved from conscription to voluntary enlistment in 1973. The ASVAB includes ten sections, four of which make up the AFQT, which is used to evaluate the general aptitude of service applicants: section 2 (arithmetic reasoning), section 3 (word knowledge), section 4 (paragraph comprehension), and half of section 5 (numerical operations). In an effort to update enlistment norms, the ASVAB was administered to the respondents of the NLSY in the summer and fall of 1979. The NLSY respondents were chosen because they formed a nationally representative sample of young people born between 1957 and 1964.

DOMESTIC SITUATION

Domestic situation has two main components—whether the respondent is married and whether there are children present in the household:

Married. Respondent's marital status is a binary variable, with a value of 1 assigned to those who were married in 1990. In general, married individuals are less likely to be in poverty.

Children. "Having children" is a binary variable, with a value of 1 indicating the presence of one or more children as members of the household. The rationale for using a binary variable is that being a parent imposes certain status and lifestyle constraints. As any parent will readily attest, the change from having no children to becoming a parent is much more momentous, from a lifestyle and standard-of-living point of view, than having a second or third child. In general, households with children are more likely to be in poverty than households without children.

NOTES

INTRODUCTION

1. Multiple configurations of causally relevant conditions may be linked to a given outcome. If there are multiple configurations, it is usually possible to distinguish subtypes of the focal outcome, with each subtype matched to a different causal configuration (see chapter 2).

2. As explained in chapter 1, "classic AI" researchers are less concerned about disconfirming cases in which the antecedent conditions are present but the outcome is absent.

3. A brief overview of qualitative comparative analysis is presented in appendix A.

4. It is important to note that fuzzy-set conditions also can be transformed into "contributing-versus-irrelevant" conditions.

1. CLASSIC ANALYTIC INDUCTION

1. In this work, I refer to instances of the presence of an outcome (e.g., college graduate) as *positive* cases, while instances of the opposing category are referred to as *negative* cases (e.g., not a college graduate). This usage of *positive* versus *negative* cases should not be confused with an alternative convention that uses the same binary to differentiate cases that are theory-confirming from those that are theory-disconfirming.

2. Common-pool resources are natural resources that are used in common by many individuals, such as fisheries or irrigation systems.

3. From the perspective of King, Keohane, and Verba's *Designing Social Inquiry* (1994), Ostrom's Nobel Prize–winning research is flawed, for she is guilty of "selecting on the dependent variable."

4. Today, correlation is almost hegemonic in the practice of quantitative social science. A matrix of bivariate correlations along with the means and standard deviations of the variables is all that is required to conduct analyses using the most commonly applied technique,

multiple regression analysis, as well as most sophisticated techniques (e.g., structural equation models, or SEM; see Bollen 1989).

5. Ragin (2008) examines these relationships in terms of the *degree* to which a set-theoretic connection is consistent with necessity or sufficiency.

6. Some social scientists hesitate to use the concepts of *sufficiency* or *necessity*. For "necessary condition" they should substitute the phrase "shared antecedent condition," as in "cases with the outcome share these antecedent conditions." For "sufficient condition" they should substitute the phrase "shared outcome," as in "cases exhibiting these conditions share this outcome."

7. Strategies for addressing cell *a* cases are discussed in detail in chapter 2.

2. RECONCILING DISCONFIRMING CASES

1. For the sake of clarity and simplicity, my examples involve a single causal condition. The demonstrations hold for situations that include multiple antecedent conditions (i.e., causal recipes) as well.

2. Cells *c* and *d* are understood as instances of different outcomes and are treated separately.

3. Of course, given disconfirming cases, a researcher could conclude simply that the hypothesized condition is not a relevant antecedent.

4. THE USES OF "NEGATIVE" CASES IN SOCIAL RESEARCH

1. A brief overview of qualitative comparative analysis (QCA) is presented in appendix A.

2. Actually, it is difficult to come up with a sociological variable that is fully and empirically binary. Even something like "straight versus gay" really means "straight versus not straight." Although gender is typically treated as empirically binary (female/male), it is often contentious to do so.

3. The *complement* of a set, denoted ~A, is the set of all elements in the given universal set U that are not in set A.

4. The researcher could limit the analysis to Republican versus Democratic voters.

5. Of course, there are statistical techniques such as multinomial logit designed for multichotomous dependent variables.

6. An alternate but mathematically equivalent approach codes the reference category as −1.

7. In fact, the analysis of the negation of the outcome is considered by some a QCA "best practice" (Schneider and Wagemann 2012: 279–80).

8. The formula for the fuzzy subset relation is $\sum \min(X_i, Y_i)/\sum(X_i)$, where X_i is the degree of membership in a causal condition (or combination of conditions) and Y_i is the degree of membership in the outcome.

9. As explained in chapter 1, there are very good reasons AI researchers might expect to find an empty cell *d*. Two important factors are (1) that as more antecedent conditions are identified by the researcher and added to the mix, the number of cell *d* cases that meet them all may be correspondingly diminished; and (2) that AI, especially classic AI, tends to favor constitutive causal conditions, yielding an integral connection between antecedent conditions and the outcome.

6. THE INTERPRETIVE LOGIC OF GENERALIZED ANALYTIC INDUCTION

1. QCA uses fuzzy sets (Zadeh 1965, 1972) to operationalize conditions that vary by level or degree. See appendix B and chapter 9.

2. Often, it is prudent to set a frequency threshold to differentiate well-populated rows from less populated rows. Less populated rows are then treated as remainders, along with the rows that lack cases altogether.

3. Some authors (e.g., Schneider and Wagemann 2012) prefer the label "conservative" to the label "complex." To maintain consistency with Ragin (2008), I use *complex*.

4. Consult Ragin (1987) for a detailed exposition of incremental elimination.

5. As demonstrated in chapter 8 and appendix C, generalized AI can accommodate conditions that must be present in some contexts and absent in other contexts for the outcome to occur.

7. GENERALIZED ANALYTIC INDUCTION: A STEP-BY-STEP GUIDE

1. A researcher trained in quantitative methods would probably take a random sample of Olympic-caliber athletes and assess variation in their longevity as Olympic athletes, using longevity as a dependent variable indicative of commitment. In this approach, conditions that sustain commitment and conditions that undermine commitment would be incorporated into a single model. Note, however, that longevity as a dependent variable fails to capture sustained commitment as a qualitative accomplishment.

2. A more detailed example, along with screenshots of the use of the software, is presented in appendix C.

3. See appendix B on the use of fuzzy sets.

4. As noted in chapter 8 and illustrated in appendix C, it is not necessary with generalized AI to convert all conditions to "contributing versus irrelevant." If, for example, a condition is thought to contribute in some contexts when coded "present" and in other contexts when coded "absent," the condition can be included in the analysis as a conventional presence/absence dichotomy.

5. Step 8 is optional. The default frequency threshold is one case.

9. APPLYING GENERALIZED AI TO CONVENTIONAL QUANTITATIVE DATA

1. The negation of membership in set A is 1 – (membership in set A): $\sim A = 1 - A$, where the tilde indicates set negation. A respondent with 0.2 membership in the set with low-income parents has a membership of 0.8 in the set with not-low-income parents.

2. *Favorable domestic situation* combines the effects of the two dichotomous independent variables used in the logistic regression analysis—married versus not married and having children versus not having children.

3. Using a fuzzy set as a dependent variable entails a floor observed value of 0 and a ceiling observed value of 1.0, which could lead to nonsensical predicted values using OLS regression (i.e., predicted values greater than 1.0 or less than 0). To address this issue, I reestimated the regression using a generalized linear model with a logit link and the binomial family. The results were entirely consistent with the OLS regression; all four independent variables had significant negative effects ($p < .001$).

4. The third fuzzy-set benchmark is the crossover point—a membership score of 0.5. Because it represents maximum ambiguity in whether a case is more in or more out of the set in question, it is not used as a meaningful qualitative breakpoint in the application of generalized AI that follows.

5. For comparison purposes, I computed a logistic regression analysis using an income-to-poverty ratio of 3.0 (or less) as the cutoff for $y = 1$. The results were very close to the logistic regression analysis using an income-to-poverty ratio of 1.0 as the cutoff value. The overlaps in the confidence intervals of the regression coefficients across the two analyses were substantial. Thus, the substantive results associated with the two applications of generalized AI cannot be duplicated by varying the cutoff value in the logistic regression.

APPENDIX C. USING FSQCA SOFTWARE TO IMPLEMENT GENERALIZED AI

1. Gamson's "negative" cases were challenging groups that failed to win new advantages for their constituents, due to a wide variety of obstacles and shortcomings.

2. Another motivation for allowing lowstatus to remain uncoded could just as well be an interest in modeling the differences between SMOs representing low-status groups versus SMOs representing moderate- or high-status groups.

3. Of course, it is always possible to use multiple frequency thresholds and assess the impact of being more versus less inclusive.

REFERENCES

Athens, Lonnie. 2006. Negative Case Analysis. In Blackwell Encyclopedia of Sociology, edited by George Ritzer, 3166–70. Oxford: Blackwell.

Bansal, Pratima, and Kendall Roth. 2000. Why Companies Go Green: A Model of Ecological Responsiveness. Academy of Management Journal 43(4): 717–36.

Becker, Howard S. 1953. Becoming a Marihuana User. American Journal of Sociology 59(3): 235–42.

———. 1958. Problems of Inference and Proof in Participant Observation. American Sociological Review 23: 652–60.

———. 1998. Tricks of the Trade. Chicago: University of Chicago Press.

———. 2015. Becoming a Marihuana User. Chicago: University of Chicago Press.

Bernard, H. Russell, Amber Wutich, and Gery Ryan. 2017. Analyzing Qualitative Data: Systematic Approaches, 2nd edition. Thousand Oaks, CA: Sage.

Bloor, Michael. 1978. On the Analysis of Observational Data: A Discussion of the Worth and Uses of Inductive Techniques and Respondent Validation. Sociology 12: 542–52.

Bollen, Kenneth. 1989. Structural Equations with Latent Variables. New York: Wiley Interscience.

Braumoeller, Bear, and Gary Goertz. 2000. The Methodology of Necessary Conditions. American Journal of Political Science 44: 844–58.

Brinton, Crane. 1938. The Anatomy of Revolution. New York: Vintage.

Chirot, Daniel. 1996. Modern Tyrants. Princeton, NJ: Princeton University Press.

Cressey, Donald R. 1953. Other People's Money: A Study in the Social Psychology of Embezzlement. Glencoe, IL: Free Press.

———. 1973. Other People's Money: A Study in the Social Psychology of Embezzlement, 2nd edition. Montclair, NJ: Patterson Smith.

Denzin, Norman K. 2006. Analytic Induction. In Blackwell Encyclopedia of Sociology, edited by George Ritzer, 134–35. Oxford: Blackwell.

Diesing, Paul. 1971. Patterns of Discovery in the Social Sciences. Chicago: Aldine.

Dion, Douglas. 1998. Evidence and Inference in the Comparative Case Study. Comparative Politics 30: 127–45.

Downton, James, and Paul Wehr. 1998. Persistent Pacifism: How Activist Commitment Is Developed and Sustained. Journal of Peace Research 35(5): 531–50.

Driscoll, Daniel. 2018. Beyond Organizational Ties: Foundations of Persistent Commitment in Environmental Activism. Social Movement Studies 17(6): 697–715.

Gamson, William. 1990. The Strategy of Social Protest, 2nd edition. Belmont, CA: Wadsworth.

George, Alexander L. 1979. Case Studies and Theory Development: The Method of Structured, Focused Comparison. In Diplomacy: New Approaches in History, Theory, and Policy, edited by Paul G. Lauren, 43–68. New York: Free Press.

George, Alexander L., and Andrew Bennett. 2005. Case Studies and Theory Development. Cambridge, MA: MIT Press.

Gilgun, Jane. 1995. We Shared Something Special: The Moral Discourse of Incest Perpetrators. Journal of Marriage and Family 57(2): 265–81.

Glaser, Barney, and Anselm Strauss. 1967. The Discovery of Grounded Theory: Strategies for Qualitative Research. New York: Weidenfeld and Nicholson.

Goertz, Gary. 2017. Multimethod Research, Causal Mechanisms, and Case Studies: An Integrated Approach. Princeton, NJ: Princeton University Press.

———. 2020. Social Science Concepts and Measurement. Princeton, NJ: Princeton University Press.

Goertz, Gary, and Stephan Haggard. 2022. Large-N Qualitative Analysis (LNQA): External Validity and Generalization in Case Study and Multi-method Research. Unpublished manuscript.

Goertz, Gary, and James Mahoney. 2005. Two-level Theories and Fuzzy-Set Analysis. Sociological Methods and Research 33(4): 497–538.

———. 2012. A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences. Princeton, NJ: Princeton University Press.

Goldenberg, Sheldon. 1993. Analytic Induction Revisited. Canadian Journal of Sociology 18(2): 161–76.

Hammersley, Martyn. 2010. A Historical and Comparative Note on the Relationship between Analytic Induction and Grounded Theorising. Forum: Qualitative Social Research 11(2): art. 4.

———. 2011. On Becker's Studies of Marihuana Use as an Example of Analytic Induction. Philosophy of the Social Sciences 41(4): 535–66.

Hammersley, Martyn, and Barry Cooper. 2012. Analytic Induction Versus QCA. In Challenging the Qualitative-Quantitative Divide, edited by Barry Cooper, Judith Glaesser, Roger Gomm, and Martyn Hammersley, 129–69. New York: Continuum Books.

Hicks, Alexander. 1994. Qualitative Comparative Analysis and Analytic Induction: The Case of the Emergence of the Social Security State. Sociological Methods and Research 23(1): 86–113.

Katz, Jack. 1982. Poor People's Lawyers in Transition. New Brunswick, NJ: Rutgers University Press.

———. 1983. A Theory of Qualitative Methodology. In Contemporary Field Research, edited by Robert Emerson, 127–48. Prospect Heights, IL: Waveland.

———. 2001. Analytic Induction. In International Encyclopedia of the Social and Behavioral Sciences, edited by Neil J. Smelser and Paul B. Baltes, 480–84. Oxford: Pergamon.

Kidder, Louise. 1981. Qualitative Research and Quasi-experimental Frameworks. In Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell, edited by Marilynn Brewer and Barry Collins, 226–56. San Francisco: Jossey-Bass.

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton, NJ: Princeton University Press.

Kosko, Bart. 1993. Fuzzy Thinking: The New Science of Fuzzy Logic. New York: Hyperion.

Lee, Raymond, and Nigel Fielding. 2004. Tools for Qualitative Data Analysis. In Handbook of Data Analysis, edited by Melissa Hardy and Alan Bryman, 529–46. Thousand Oaks, CA: Sage.

Lieberson, Stanley. 1985. Making It Count: The Improvement of Social Research and Theory. Berkeley: University of California Press.

Lindesmith, Alfred. 1947. Opiate Addiction. Bloomington, IN: Principia Press.

———. 1952. Two Comments on W. S. Robinson's "The Logical Structure of Analytic Induction." American Sociological Review 17(4): 492–94.

———. 1968. Addiction and Opiates. Chicago: Aldine.

Linz, Juan, and Alfred Stepan (editors). 1978. The Breakdown of Democratic Regimes. Baltimore: Johns Hopkins University Press.

Mackie, John L. 1965. Causes and Conditionals. American Philosophical Quarterly 2: 245–65.

———. 1980. The Cement of the Universe. New York: Oxford.

Mahoney, James, and Gary Goertz. 2004. The Possibility Principle: Choosing Negative Cases in Comparative Research. American Political Science Review 98: 653–69.

Manning, Peter. 1982. Analytic Induction. In Handbook of Social Science Methods: Qualitative Methods, edited by Robert Smith and Peter Manning, 273–302. Cambridge, MA: Ballinger.

Mill, John Stuart. 1967 [1843]. A System of Logic: Ratiocinative and Inductive. Toronto: University of Toronto Press.

Miller, Steven. 1982. Quality and Quantity: Another View of Analytic Induction. Quality and Quantity 16(4): 281–95.

Monaghan, Lee. 2002. Vocabularies of Motives for Illicit Steroid Use among Bodybuilders. Social Science and Medicine 55(5): 695–708.

NLSY79 User's Guide. 1999. A Guide to the 1979–1998 National Longitudinal Survey of Youth Data. Columbus, OH: Center for Human Resource Research, Ohio State University.

O'Donnell, Guillermo. 1973. Modernization and Bureaucratic-Authoritarianism. Berkeley: University of California Press.

O'Donnell, Guillermo, and Philippe Schmitter. 1986. Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies. Baltimore: Johns Hopkins University Press.

Ostrom, Elinor. 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge, MA: Cambridge University Press.

Ragin, Charles C. 1987. The Comparative Method: Moving beyond Qualitative and Quantitative Strategies. Berkeley: University of California Press.

———. 1992. Casing and the Process of Social Inquiry. In What Is a Case?, edited by Charles Ragin and Howard Becker, 217–26. New York: Cambridge University Press.

———. 1997. Turning the Tables: How Case-Oriented Methods Challenge Variable-Oriented Methods. Comparative Social Research 16: 27–42.

———. 2000. Fuzzy-Set Social Science. Chicago: University of Chicago Press.

———. 2006a. Set Relations in Social Research: Evaluating Their Consistency and Coverage. Political Analysis 14(3): 291–310.

———. 2006b. The Limitations of Net Effects Thinking. In Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide, edited by Benoit Rihoux and Heike Grimm, 13–42. New York: Springer.

———. 2008. Redesigning Social Inquiry: Fuzzy Sets and Beyond. Chicago: University of Chicago Press.

———. 2009. Reflections on Casing and Case-Oriented Research. In The Sage Handbook of Case-Based Methods, edited by David Byrne and Charles Ragin, 522–34. London: Sage.

———. 2014. The Comparative Method: Moving beyond Qualitative and Quantitative Strategies. Berkeley: University of California Press.

———. 2021. User's Guide to Fuzzy-Set/Qualitative Comparative Analysis, version 4. www.fsqca.com.

Ragin, Charles C., and Lisa Amoroso. 2019. Constructing Social Research: The Unity and Diversity of Method, 3rd edition. Thousand Oaks, CA: Pine Forge.

Ragin, Charles C., and Sean Davey. 2021. fsQCA. www.fsqca.com.

Ragin, Charles C., and Peer Fiss. 2017. Intersectional Inequality: Race, Class, Test Scores and Poverty. Chicago: University of Chicago Press.

Robinson, W. S. 1951. The Logical Structure of Analytic Induction. American Sociological Review 16: 812–18.

Schneider, Carsten, and Ingo Rohlfing. 2016. Case Studies Nested in Fuzzy-Set QCA on Sufficiency: Formalizing Case Selection and Causal Inference. Sociological Methods and Research 45(3): 526–68.

Schneider, Carsten, and Claudius Wagemann. 2012. Set-Theoretic Methods for the Social Sciences. New York: Cambridge University Press.

Skocpol, Theda. 1979. States and Social Revolutions: A Comparative Analysis of France, Russia, and China. Cambridge: Cambridge University Press.

Smithson, Michael. 1987. Fuzzy Set Analysis for the Behavioral and Social Sciences. New York: Springer.

Smithson, Michael, and Jay Verkuilen. 2006. Fuzzy Set Theory. Thousand Oaks, CA: Sage.

Stokke, Olav Schram. 2004. Boolean Analysis, Mechanisms, and the Study of Regime Effectiveness. In Regime Consequences: Methodological Challenges and Research Strategies, edited by Arild Underdal and Oran R. Young, 87–120. Dordrecht, The Netherlands: Kluwer Academic.

Tacq, Jacques. 2007. Znaniecki's Analytic Induction as a Method of Sociological Research. Polish Sociological Review 158: 187–208.

Tavory, Iddo, and Stefan Timmermans. 2014. Abductive Analysis: Theorizing Qualitative Research. Chicago: University of Chicago Press.

Turner, Ralph H. 1953. The Quest for Universals in Sociological Research. American Sociological Review 18: 604–11.

Vaughan, Diane. 1986. Uncoupling: Turning Points in Intimate Relationships. New York: Oxford University Press.

Viterna, Jocelyn. 2006. Pulled, Pushed, and Persuaded: Explaining Women's Mobilization into the Salvadoran Guerrilla Army. American Journal of Sociology 112(1): 1–45.

Walker, Henry, and Bernard Cohen. 1985. Scope Statements: Imperatives for Evaluating Theory. American Sociological Review 50: 288–301.

Weber, Max. 1949 [1905]. Objective Possibility and Adequate Causation in Historical Explanation. In The Methodology of the Social Sciences, edited by Edward A. Shils and Henry A. Finch, 164–87. Glencoe, IL: Free Press.

Zadeh, Lotfi. 1965. Fuzzy Sets. Information and Control 8: 338–53.

———. 1972. Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges. Journal of Cybernetics 2(3): 4–34.

Znaniecki, Florian. 1934. The Method of Sociology. New York: Farrar & Rinehart.

INDEX

Founded in 1893,
UNIVERSITY OF CALIFORNIA PRESS
publishes bold, progressive books and journals on topics in the arts, humanities, social sciences, and natural sciences—with a focus on social justice issues—that inspire thought and action among readers worldwide.

The UC PRESS FOUNDATION
raises funds to uphold the press's vital role as an independent, nonprofit publisher, and receives philanthropic support from a wide range of individuals and institutions—and from committed readers like you. To learn more, visit ucpress.edu/supportus.

This book explores analytic induction, an approach to the analysis of cross-case evidence on qualitative outcomes that has deep roots in sociology. A popular research technique in the early decades of empirical sociology, analytic induction differs fundamentally as a method of social research from conventional variation-based approaches. In *Analytic Induction for Social Research*, Charles C. Ragin demonstrates that much is gained from systematizing analytic induction. The approach he introduces here offers a new template for conducting cross-case analysis and provides a new set of tools for answering common research questions that existing methods cannot address.

"Breaks new ground in introducing analytic induction as an approach distinct from qualitative comparative analysis. Charles Ragin's writing is among the clearest, most accessible, and engaging that I know."
Peer C. Fiss, Jill and Frank Fertitta Chair in Business Administration and Professor of Management and Organization, University of Southern California

"At a time when methodological debates are becoming increasingly mathematical, this intervention is both refreshingly nontechnical and unusually helpful for qualitative researchers in sociology and political science. Because of its clarity, brevity, and usability, qualitative researchers in the social sciences are going to want a copy of this book."
James Mahoney, Gordon Fulcher Professor in Decision-Making and Professor of Sociology and Political Science, Northwestern University

Charles C. Ragin is Chancellor's Professor Emeritus of Sociology at the University of California, Irvine. He is a recipient of the International Science Council's Stein Rokkan Prize, the Policy Studies Organization's Donald Campbell Award, and the American Sociological Association's Paul F. Lazarsfeld Award.

UNIVERSITY OF CALIFORNIA PRESS
WWW.UCPRESS.EDU

A free ebook version of this title is available through Luminos, University of California Press's Open Access publishing program. Visit www.luminosoa.org to learn more.

Cover illustration: Courtesy of the author
Cover design: Michelle Black

ISBN: 978-0-520-39373-8
9 780520 393738