11: Net Effects versus Configurations:
An Empirical Demonstration
coauthored with Peer Fiss

This chapter presents a critique of net-effects thinking in a practical manner by contrasting a conventional net-effects analysis of a large-*N*, policy-relevant data set (the National Longitudinal Survey of Youth, or NLSY, also known as the *Bell Curve* data, from Herrnstein and Murray's 1994 publication *The Bell Curve*) with an alternate analysis of the same data, following the principles developed in this book. While the two approaches differ in several important respects, the key difference is that the net-effects approach focuses on the independent effects of causal variables on the outcome, while the configurational approach attends to combinations of causal conditions and attempts to establish explicit links between specific combinations of conditions and the outcome. This alternate method, known as fuzzy-set qualitative comparative analysis (fsQCA), combines the use of fuzzy sets with the analysis of cases as configurations, a central feature of case-oriented social research (Ragin 1987). In this approach, each case is examined in terms of its degree of membership in different *combinations* of causally relevant conditions. Using fsQCA, researchers can consider cases' varying degree of membership in all of the logically possible combinations of a given set of causal conditions and then use set-theoretic methods to analyze—in a logically disciplined manner—the varied connections between causal combinations and the outcome.

I offer this alternate approach not as a replacement for net-effects analysis but as a complementary technique. Fuzzy-set qualitative comparative analysis is best understood as an exploratory/interpretive technique, grounded in set theory. While probabilistic criteria can be incorporated into fsQCA, it is not an inferential technique, per se. It is

an alternate way of analyzing evidence, starting from very different assumptions regarding the kinds of "findings" that social scientists seek. These alternate assumptions reflect the logic and spirit of qualitative research, where investigators study cases as configurations, with an eye toward how the different parts or aspects of cases fit together.

A Net-Effects Analysis of the Bell Curve Data

In *The Bell Curve,* Herrnstein and Murray (1994) compute rudimentary logistic regression analyses to gauge the importance of Armed Forces Qualification Test (AFQT) scores on a variety of outcomes. They control for the effects of only two competing variables in most of their main analyses, respondent's age (at the time the AFQT was administered) and parental socioeconomic status (SES). Their central finding is that AFQT score (which they interpret as a measure of general intelligence) is more important than parental SES when considering major life outcomes such as avoiding poverty. They interpret this and related findings as proof that, in modern society, "intelligence" (which they assert is inborn) has become the most important factor shaping life chances. Their explanation focuses on the fact that the nature of work has changed and that today a much higher labor market premium is attached to high cognitive ability.

Herrnstein and Murray's main findings with presence/absence of poverty as the outcome of interest are presented in table 11.1 (with absence of poverty = 1). The reported analysis uses standardized data (*z* scores) for both parental SES and AFQT score to facilitate comparison of effects. The analysis shown in table 11.1 is limited to black males with complete data on all the variables used in this analysis and in the subsequent analyses reported in this chapter, including the fuzzy-set analysis. The strong positive impact of AFQT scores, despite the statistical control for the effect of parental SES, mirrors the *Bell Curve* results.

A major rebuttal of the *Bell Curve* "thesis," as it became known, was presented by a team of University of California at Berkeley sociologists, Claude Fischer, Michael Hout, Martin Sanchez Jankowsk,

Table 11.1: Logistic regression of poverty avoidance on AFQT scores, parental SES, and age (Bell Curve model; black males only)

	B	S.E.	Sig.	Exp(B)
AFQT (*z* score)	0.651	0.139	0.000	1.917
Parental SES (*z* score)	0.376	0.117	0.001	1.457
Age	0.040	0.050	0.630	1.040
Constant	1.123	0.859	0.191	3.074

Note: Chi-squared = 53.973, df = 3. B = regression coefficient, S.E. = standard error of regression coefficient, Sig. = statistical significance.

Samuel Lucas, Ann Swidler, and Kim Voss (1996). In their book, *Inequality By Design,* they present a much more elaborate logistic regression analysis of the NLSY data. Step by step, they include more and more causal conditions (e.g., neighborhood and school characteristics) that they argue should be seen as competitors with AFQT scores. In their view, AFQT score has a substantial effect in the *Bell Curve* analysis only because the logistic regression analyses that Herrnstein and Murray report are radically underspecified. To remedy this problem, Fischer et al. include more than fifteen control variables in their analysis of the effects of AFQT scores on the odds of avoiding poverty. While this "everything but the kitchen sink" approach dramatically reduces the impact of AFQT scores on poverty, the authors leave themselves open to the charge that they have misspecified their analyses by being overinclusive.

Table 11.2 reports the results of a logistic regression analysis of poverty using only a moderate number of independent variables. Specifically, presence/absence of poverty (with absence = 1) is regressed on five independent variables: AFQT score, years of education, parental income, married versus not married, and one or more children versus no children. The three interval-scale variables are standardized (using *z* scores) to simplify the comparison of effects. Like the previous analysis, the table shows the results for black males only. The rationale for this specification, using five independent variables, is that the model is more fully specified than the radically spare model presented by Herrnstein and Murray and less elaborate and cumbersome

than Fischer et al.'s kitchen-sink model. In other words, this logistic regression analysis attempts to strike a balance between the two specification extremes, while focusing on several of the most important causal conditions. The results presented in table 11.2 are consistent with both Herrnstein and Murray and Fischer et al. in that they show that AFQT score has an independent impact on poverty avoidance, but not nearly as strong as that reported by Herrnstein and Murray. Consistent with Fischer et al., table 11.2 shows very strong effects of competing causal conditions, especially years of education and marital status. These conditions were not included in the *Bell Curve* analysis.

More generally, table 11.2 confirms the specification dependence of net-effects analysis. For example, if years of education is "accepted" as a competing cause (and not considered derivative of AFQT scores), then it is clearly more important than test scores. Likewise, the impact of marriage on the odds of staying out of poverty is substantial for black males. According to table 11.2, married black males are more than five times more likely to avoid poverty that unmarried black males. Even though these results, like all net-effects analyses, are specification dependent, the fact that a very modest number of competing independent variables greatly reduces the estimate of the effect of AFQT scores on poverty casts substantial doubt on the *Bell Curve* thesis.

Table 11.2: Logistic regression of poverty avoidance on AFQT scores, parental income, years of education, marital status, and children (black male sample)

	B	S.E.	Sig.	Exp(B)
AFQT (*z* score)	0.391	0.154	0.011	1.479
Parental income (*z* score)	0.357	0.154	0.020	1.429
Education (*z* score)	0.635	0.139	0.000	1.887
Married (yes = 1, 0 = no)	1.658	0.346	0.000	5.251
Children (yes = 1, 0 = no)	−0.524	0.282	0.063	0.592
Constant	1.970	0.880	0.025	7.173

Note: Chi-squared = 104.729, df = 5. B = regression coefficient, S.E. = standard error of regression coefficient, Sig. = statistical significance.

A Reanalysis of the *Bell Curve* Data Using fsQCA

The success of any fuzzy-set analysis depends on the careful construction and calibration of the fuzzy sets. The core of both crisp-set and fuzzy-set analysis is the evaluation of set-theoretic relationships, for example, the assessment of whether membership in a combination of causal conditions can be considered a consistent subset of membership in a given outcome. A fuzzy subset relationship exists when the scores in one set (e.g., the fuzzy set of individuals who combine high parental income, college education, high test scores, and so on) are consistently less than or equal to the scores in another set (e.g., the fuzzy set of individuals not in poverty). Thus, it matters a great deal how fuzzy sets are constructed and how membership scores are calibrated. Serious miscalibrations can distort or undermine the identification of set-theoretic relationships. By contrast, for the conventional variable to be useful in a net-effects analysis, it needs only to vary in a meaningful way (see chapter 4). Often, the specific metric of a conventional variable is ignored altogether by researchers because it is arbitrary or meaningless. Even when a variable has a meaningful metric, researchers often focus only on the direction and significance of its effect.

In order to calibrate fuzzy-set membership scores, researchers must use their substantive knowledge (see chapter 5). The resulting membership scores must have face validity in relationship to the set in question, especially how it is conceptualized and labeled. A fuzzy score of 0.25, for example, has a very specific meaning—that a case is halfway between "full exclusion" from a set (e.g., a membership score of 0.0 in the set of individuals with high parental income) and the crossover point (0.5, the point of maximum ambiguity in whether a case is more in or more out of this set). As explained in *Fuzzy-Set Social Science* (Ragin 2000) and chapters 4 and 5 of this book, the most important decisions in the calibration of a fuzzy set involve the definition of the three qualitative anchors that structure a fuzzy set: full inclusion in the set, the crossover point (membership = 0.5), and full exclusion from the set. The main sets used in the analysis reported in this chapter are

degree of membership in the outcome—the set of individuals avoiding poverty—and degree of membership in sets reflecting five background characteristics: parental income, AFQT scores, education, marital status, and children. The calibration of these fuzzy sets is detailed in the practical appendix at the end of this chapter.

At this point it is important to note that representing a single interval-scale variable with two fuzzy sets is often fruitful. For example, the variable parental income can be transformed separately into the set of individuals with high parental income and the set of individuals with low parental income. It is necessary to construct *two* fuzzy sets because of the *asymmetry* of the two target concepts. Full *non*membership in the set of individuals with high parental income (a membership score of 0.0 in high parental income) does *not* imply full *membership* in the set with low parental income (a score of 1.0), for it is possible to be fully out of the set of individuals with high parental income without being fully in the set of individuals with low parental income. The same is true for the other two interval-scale variables used as causal conditions in the logistic regression analysis (table 11.2), AFQT scores and years of education. This dual coding of key causal conditions has important theoretical benefit. For example, is it having a *high* AFQT score that is linked to superior life chances, or is it *not* having a *low* AFQT score that matters? This issue is especially important because Herrnstein and Murray (1994) argue that having a high AFQT score (which they interpret as having high cognitive ability) is the key to success in modern society.

Note also that the language and logic of "variables" does not translate directly into set theory. A case cannot have membership in a variable, for example, a high degree of membership in AFQT or a high degree of membership in parents' income. Instead a case has membership in a set, for example, strong membership in the set of people with high AFQT scores or strong membership in the set of people with high parental income. The translation of variables to sets requires careful definition and labeling of the target sets, which in turn provides the primary basis for calibrating membership. Thus, when translating such variables as parental income to fuzzy sets, it is useful to consider the different target sets that can be created from a single source variable, especially in light of the theoretical and substantive issues that inspire and guide the research.

Altogether, the fuzzy-set analysis reported in this chapter uses eight causal conditions. Two are crisp sets: married versus not married and one or more children versus no children. The remaining six are fuzzy sets: degree of membership in the set of cases with high parental income, degree of membership in the set of cases with low parental income, degree of membership in the set of cases with high AFQT scores, degree of membership in the set of cases with low AFQT scores, degree of membership in the set of cases with college education, and degree of membership in the set of cases with high school education.

After calibrating the fuzzy sets, the next task is to calculate the degree of membership of each case in each of the 2^k logically possible combinations of eight causal conditions, and then to assess the distribution of cases across these combinations. With eight causal conditions, there are 256 logically possible combinations of conditions.[1] Table 11.3 lists the 42 of these 256 combinations that have at least four cases with greater than 0.5 membership.[2] Recall that a case can have, at most, only one configuration membership score that is greater than 0.5. Thus, the 256 combinations of conditions can be evaluated with respect to case frequency by examining the number of empirical instances of each combination. If a configuration has no cases with greater than 0.5 membership, then there are no cases that are more in than out of the set defined by the combination of conditions (and no cases in the corresponding sector of the multidimensional vector space defined by the causal conditions).

Table 11.3 reveals that the data used in this analysis (and, by implication, in the logistic regression analyses reported in tables 11.1

1. Of course, many of these 256 combinations are not empirically possible. For example, a case cannot have high membership in both high income parents and low income parents. The number of empirically possible combinations is 108. This number still dwarfs the number of high-frequency combinations (see table 11.3).

2. An additional nineteen rows (not shown in table 11.3) have one, two, or three cases each. The remaining rows have no cases with greater than 0.5 membership.

and 11.2) are remarkably limited in their diversity. Altogether, only 42 of the 256 sectors contained within the eight-dimensional vector space have at least four empirical instances (i.e., at least four cases with greater then 0.5 membership in the corner), and most of the frequencies reported in the table are quite small. The two most populated sectors capture 25 percent of the cases; the seven most populated capture half of the cases; and the fourteen most populated capture nearly 70 percent of the cases. The number of well-populated sectors (fourteen) is small even relative to the number of sectors that exist in a five-dimensional vector space ($2^5 = 32$). This is the number of sectors that would have been obtained if the three interval-level variables (years of education, parental income, and AFQT scores) used in the logistic regression analysis had been transformed into one fuzzy set each instead of two.

In fuzzy-set analyses of this type (large *N*), it is important to establish a strength-of-evidence threshold for combinations of conditions, using the information on the distribution of strong instances across sectors. Specifically, causal combinations with only a few strong instances (i.e., a few cases with greater than 0.5 membership in the combination) should be filtered out and not subject to further empirical analysis. In addition to the fact that it would be unwise to base a conclusion about a combination of individual-level attributes on a small number of instances, the existence of cases in low-frequency sectors may be due to measurement or assignment error. The fuzzy-set analysis that follows uses a frequency threshold of at least ten strong instances. This value was selected because it captures more than 80 percent of the cases assigned to combinations. Using this rule, the twenty-three most common combinations of conditions are retained in this analysis. The low-frequency rows (including those shown in the bottom part of table 11.3 with frequencies ranging from four to nine) are filtered out of the analysis. Because these rows do not meet the strength-of-evidence threshold, they are treated as "remainder" combinations in the analysis that follows.

The next task is to assess the consistency of the evidence for each of the combinations of conditions (the twenty-three high-frequency

Table 11.3: Distribution of cases across vector space corners (the 42 combinations with at least 4 cases each)

Married	Children	High parental income	Low parental income	High AFQT score	Low AFQT score	High school educated	College educated	Count	Cumulative proportion
0	0	0	1	0	1	1	0	118	0.152
0	0	0	0	0	1	1	0	78	0.253
0	0	0	0	0	0	1	0	53	0.321
1	1	0	0	0	1	1	0	41	0.375
1	1	0	0	0	0	1	0	39	0.425
1	1	0	1	0	1	1	0	34	0.469
0	0	0	1	0	0	1	0	30	0.508
0	0	0	0	0	0	1	1	23	0.537
0	1	0	0	0	1	1	0	22	0.566
0	0	1	0	0	1	1	0	20	0.592
0	1	0	1	0	1	1	0	20	0.618
0	0	1	0	0	0	1	0	19	0.642
1	1	0	1	0	0	1	0	19	0.667
0	0	0	1	0	1	0	0	18	0.690
0	0	0	1	0	0	1	1	12	0.705
0	0	1	0	0	0	1	1	12	0.721
0	0	0	0	0	1	0	0	11	0.735
0	1	0	0	0	0	1	0	11	0.749
1	1	1	0	0	0	1	0	11	0.764
1	1	1	0	0	1	1	0	11	0.778
0	0	0	1	0	1	1	1	10	0.791
1	0	0	1	0	1	1	0	10	0.804
1	1	0	0	0	0	1	1	10	0.817
1	0	0	1	0	0	1	0	9	0.828
1	0	1	0	0	0	1	1	9	0.840
1	1	1	0	0	0	1	1	9	0.851
1	0	0	0	0	0	1	0	7	0.860
1	0	0	0	0	1	1	0	7	0.870
0	0	0	0	0	1	1	1	6	0.877
0	1	0	1	0	1	0	0	6	0.885
1	0	1	0	0	0	1	0	6	0.893
1	0	1	0	0	1	1	0	6	0.901
1	1	0	0	0	1	0	0	6	0.908
0	0	1	0	1	0	1	1	5	0.915
1	1	0	1	0	0	1	1	5	0.921
0	0	0	0	0	0	0	0	4	0.926
0	0	1	0	0	1	1	1	4	0.932
0	1	0	0	0	0	1	1	4	0.937
0	1	0	1	0	0	1	0	4	0.942
1	0	0	0	0	0	1	1	4	0.947
1	1	0	0	0	1	1	1	4	0.952
1	1	0	1	0	1	0	0	4	0.957
All remaining combinations of conditions; row counts < 4									1.000

rows from table 11.3) with the subset relation. Specifically, it is necessary to determine whether degree of membership in each combination of conditions is a subset of degree of membership in the outcome. As explained in chapter 1, the subset relation is used to assess causal sufficiency. With fuzzy sets, the subset relation is demonstrated by showing that degree of membership in a combination of conditions (which can range from 0.0 to 1.0) is consistently less than or equal to degree of membership in the outcome. These assessments use all cases in each assessment, including cases with less than 0.5 membership in a given combination. Such cases may be inconsistent, and their inconsistency counts against the set-theoretic relation in question. For example, a case with a membership of 0.40 in a causal combination and a membership of 0.20 in the outcome would lower the consistency score for that combination, even though this case is more out than in both the combination and the outcome.

As shown in chapter 3, a simple descriptive measure of the degree to which the evidence regarding a combination of conditions is consistent with the subset relation with respect to the outcome is:

$$\Sigma[\min(\mathbf{X}_i,\mathbf{Y}_i)]/\Sigma(\mathbf{X}_i)$$

where min indicates selection of the lower of the two scores, $\mathbf{X}_i$ indicates degree of membership in a combination of conditions, and $\mathbf{Y}_i$ indicates degree of membership in the outcome. When all $\mathbf{X}_i$ values are consistent (i.e., their membership scores in the combination are uniformly less than or equal to their corresponding $\mathbf{Y}_i$ values), the calculation yields a score of 1.0. If many of the $\mathbf{X}_i$ values exceed their $\mathbf{Y}_i$ values by a substantial margin, however, the resulting score is substantially less than 1.0. Generally, scores on this measure that are lower than 0.75 indicate substantial departure from the set-theoretic relation $\mathbf{X}_i \leq \mathbf{Y}_i$.

Table 11.4 reports the results of the set-theoretic consistency assessments for the twenty-three combinations in table 11.3 that meet the strength-of-evidence threshold (a frequency of at least ten cases that are more in than out of each combination). The consistency scores for the combinations range from 0.340 to 0.986, indicating a substantial spread in the degree to which the subset relation is satisfied. In the truth table analysis that follows, the seven combinations with consistency

scores of at least 0.80 are treated as subsets of the outcome; the remaining sixteen fail to satisfy this criterion. Once this distinction is made, table 11.4 can be analyzed as a truth table (see chapter 7). The binary outcome, which is based on the fuzzy set–theoretic consistency scores in the adjacent column, is listed in the last column of table 11.4.

Using fsQCA (Ragin, Drass, and Davey 2007), it is possible to derive two truth table solutions, one maximizing parsimony and the other maximizing complexity (see chapter 9). The most parsimonious solution permits the incorporation of *any* counterfactual combination that contributes to the derivation of a logically simpler solution. This solution of the truth table yields three relatively simple combinations linked to poverty avoidance:

married·~children +
high_income·~low_AFQT +
college·~low_AFQT

where (here and in subsequent fsQCA results) college is the fuzzy set for college educated, high_school is the fuzzy set for high school educated, low_income is the fuzzy set for low parental income, high_income is the fuzzy set for high parental income, low_AFQT is the fuzzy set for low AFQT score, high_AFQT is the fuzzy set for high AFQT score, children is the crisp set for at least one child, married is the crisp set for married, ~ indicates negation or "not," · signals combined conditions (set intersection), and + signals alternate combinations of conditions (set union). The parsimonious solution reveals that the three combinations of conditions linked to poverty avoidance are (1) being married combined with not having children, (2) having high income parents combined with not having a low AFQT score, and (3) having a college degree combined with not having a low AFQT score.

While parsimonious, this solution incorporates many counterfactual combinations (i.e., remainders), and many of these, in turn, are "difficult" from the perspective of existing theoretical and substantive knowledge (see chapters 8 and 9). For example, the combination of not being high school educated but being married and not having children is included in the first combination listed above. Too few empirical instances of this combination are present to allow its assessment, but

Table 11.4: Assessments of set-theoretic consistency (for the 23 configurations passing frequency threshold of at least 10 cases)

Married	Children	High parental income	Low parental income	High AFQT score	Low AFQT score	High school	College	Count	Consistency	Outcome
0	0	1	0	0	0	1	1	12	0.986	1
1	1	0	0	0	0	1	1	10	0.893	1
0	0	0	1	0	0	1	1	12	0.892	1
1	1	1	0	0	0	1	0	11	0.884	1
1	0	0	1	0	1	1	0	10	0.876	1
0	0	0	0	0	0	1	1	23	0.864	1
0	0	1	0	0	0	1	0	19	0.830	1
1	1	1	0	0	1	1	0	11	0.792	0
0	0	0	0	0	0	1	0	53	0.788	0
0	0	0	1	0	1	1	1	10	0.767	0
1	1	0	0	0	0	1	0	39	0.754	0
1	1	0	0	0	1	1	0	41	0.706	0
1	1	0	1	0	1	1	0	34	0.657	0
1	1	0	1	0	0	1	0	19	0.641	0
0	0	0	0	0	1	1	0	78	0.636	0
0	0	1	0	0	1	1	0	20	0.620	0
0	0	0	1	0	0	1	0	30	0.617	0
0	1	0	0	0	0	1	0	11	0.578	0
0	0	0	0	0	1	0	0	11	0.498	0
0	0	0	1	0	1	1	0	118	0.482	0
0	1	0	0	0	1	1	0	22	0.402	0
0	1	0	1	0	1	1	0	20	0.376	0
0	0	0	1	0	1	0	0	18	0.340	0

the parsimonious solution assumes that individuals with this combination are able to avoid poverty, despite their failure to complete high school. With 256 logically possible combinations of conditions, many combinations are without cases or with very few cases, as table 11.3 indicates. The parsimonious solution just presented incorporates many such combinations, without regard for their empirical plausibility—that is, without regard for existing substantive knowledge.

If, instead, the researcher evaluates the plausibility of the counterfactual combinations, a less parsimonious ("intermediate") solution can be derived.[3] This intermediate solution is obtained by first deriving the most complex solution (not shown here) and then using only "easy" counterfactuals to produce an intermediate solution, as explained in chapter 9.[4] The intermediate solution is a subset of the most parsimonious solution and a superset of the most complex solution.

The intermediate solution indicates that five combinations of conditions are linked to poverty avoidance:

married·~children·high_school +
married·high_income·~low_AFQT·high_school +
~children·high_income·~low_AFQT·high_school +
~children·~low_AFQT·college +
married·~low_income·~low_AFQT·college

These five combinations linked to poverty avoidance are similar in that they all include education (college or high_school) and some aspect of household composition (married or ~children or both). Four include not having low AFQT scores (~low_AFQT) as an ingredient, and four include conditions related to parental income (either high

3. The software package fsQCA will produce all three solutions (complex, parsimonious, and intermediate) when the Standard Analysis button is clicked at the bottom of the truth table spreadsheet. The user is then prompted for the input that is the basis for the derivation of the intermediate solution.

4. The substantive knowledge that is incorporated into the production of the intermediate solution in the present analysis is quite simple. For example, it is assumed that having a high school education (as opposed to not having completed high school) is linked to staying out of poverty, that having parents who are not low income is linked to staying out of poverty, that being married is linked to staying out of poverty, and so on.

parental income or not-low parental income). These results are important because they confirm that the causal conditions linked to poverty avoidance are combinatorial in nature and that it is possible to discern the relevant combinations when cases are viewed as configurations.

Recall from chapter 9 that the terms included in the parsimonious solution *must* be included in *any* representation of the results, for these are the decisive causal ingredients that distinguish combinations of conditions that are consistent subsets of the outcome from those that are not (that is, among the combinations that pass the frequency threshold). Thus, these ingredients should be considered the "core" causal conditions. The ingredients that are added in the intermediate solution are those that are also present in the cases that consistently display the outcome but that require difficult counterfactuals to remove. Thus, these conditions are "complementary" or "contributing" conditions in the sense that they make sense as important contributing factors and can be removed from the solution only if the researcher is willing to make assumptions that are at odds with existing substantive and theoretical knowledge. This researcher might have to assume, for example, that a high school dropout with a given set of characteristics (e.g., married without children) would be able to avoid poverty. Table 11.5 summarizes the five solutions in a way that differentiates core versus complementary causal conditions. This table also reports the consistency, raw coverage, and unique coverage calculations for each of the five recipes. (These calculations are explained in chapter 3.)

The results also can be summarized with the aid of a table that sorts the different recipes for poverty avoidance according to the respondent's family status. Table 11.6 shows that different recipes are clearly evident for black males in different family status categories. Those who are married and without children have the easiest time avoiding poverty. All that is required is a high school education. At the other extreme, there are no recipes for poverty avoidance for black males who are unmarried with children. For black males who are unmarried and without children, the recipe is to combine not-low AFQT scores with either college education or high school education combined with

Table 11.5: Configurations for avoiding poverty for black males

	Solution				
	1	2	3	4	5
Family Status					
Married	●	•			•
Children	⊖		ө	ө	
Education					
High school	•	•	•		
College				●	●
Test Scores					
High AFQT					
Low AFQT		⊖	⊖	⊖	⊖
Parental Income					
High income		●	●		
Low income					ө
Consistency	0.92	0.94	0.91	0.92	0.95
Raw coverage	0.13	0.10	0.14	0.16	0.11
Unique coverage	0.07	0.02	0.04	0.06	0.03

Note: ● = core causal condition (present); ⊖ = core causal condition (absent); • = contributing causal condition (present); ө = contributing causal condition (absent).

Table 11.6: Results of fuzzy-set analysis sorted according to family status

Family status	Recipe for poverty avoidance
Married, no children	high_school
Unmarried, no children	~low_AFQT·(college + high_school·high_income)
Married, children	~low_AFQT·(college·~low_income + high_school·high_income)
Unmarried, children	{Ø}

high-income parents. For black males who are married with children, the recipe is similar, but slightly more complex: they combine not-low AFQT scores with either college education and not-low-income parents or high school education and high-income parents. In short, the table shows that domestic situation has a very powerful impact on the resources that are required for avoiding poverty.

In addition to revealing the combinatorial complexities of staying out of poverty for black males, the results also challenge the interpretation of AFQT scores offered by Herrnstein and Murray (1994). Recall that the core of their argument is that the nature of work has changed and that the labor market now places a premium on high cognitive ability. The image they conjure is one of a society that has many positions for the cognitively gifted but fewer slots for those who are more modest in their cognitive endowments. The results presented here are unequivocal: what really matters when it comes to avoiding poverty is to *not* have *low* test scores. In other words, following Herrnstein and Murray's argument, one would expect high cognitive skills to be a common ingredient in these solutions; instead, it is clear that the cognitive bar is much lower. The key is to *not* have low cognitive ability, which indicates in turn that modest cognitive ability remains adequate in today's world. Of course, this interpretation assumes that one accept the questionable claim that AFQT scores indicate cognitive ability. According to many of the critics of the *Bell Curve* thesis, AFQT scores indicate the acquisition of cultural capital. In this light, the findings reported here indicate that one ingredient in the effort to avoid poverty is the possession of at least modest cultural capital.

Discussion

The results presented here are preliminary findings drawn from a larger fuzzy-set analysis of the *Bell Curve* data. The primary goal of this illustrative research is to provide a contrast between a net-effects analysis and a configurational analysis of the same data.

The contrast between the two approaches is clear. The findings of the net-effects analysis are expressed in terms of separate variables.

They provide the final tally in the competition to explain variation in the outcome, avoiding poverty. Education and marital status win this competition, but AFQT is not eliminated, for it retains a modest net effect, despite stiff competition (compare table 11.1 and table 11.2). The logistic regression results are silent on the issue of causal combinations; the analysis of causal combinations would require the examination of complex interaction models. Examining a saturated interaction model, for example, would require the estimation of thirty-two coefficients in a single equation. Even if such a model could be estimated (extreme collinearity makes this task infeasible), the model would be virtually impossible to interpret, once estimated.

Note also that the assumptions of additivity and linearity in the logistic regression analysis allow the estimation of outcome probabilities for all thirty-two sectors of the vector space defined by the five independent variables, regardless of whether these sectors are populated with cases. Thus, the net-effects approach addresses the problem of limited diversity in an indirect and covert manner by assuming that the effect of a given variable is the same regardless of the values of the other variables and that a linear relationship can be extrapolated beyond an observed range of values. To derive the estimated probability of avoiding poverty for any point in the vector space defined by the independent variables, it is necessary simply to insert the coordinates of that point into the equation and calculate the predicted value. The issue of limited diversity is thus sidestepped altogether.

By contrast, this issue must be confronted head-on in a configurational analysis. Naturally occurring data are profoundly limited in their diversity, as illustrated in table 11.3. This fact is apparent whenever researchers examine the distribution of cases across logically possible combinations of conditions, especially when the number of conditions is more than a few. As the analysis reported here illustrates, the problem of limited diversity is *not* remedied by having a large number of cases.

When cases are viewed configurationally, it is possible to identify the different combinations of conditions linked to an outcome. The results of the configurational analyses reported in this chapter show that there are several recipes for staying out of poverty for black males in the United States. The recipes all include educational qualifications of some sort (high school or college) and a favorable household composition (either marriage or being childless or both). Not having low AFQT scores is also a condition in four of the five causal recipes, as is having either high or not-low parental income, in these same four recipes. Herrnstein and Murray (1994) dramatize the implications of their research by claiming that if one could choose at birth between having a high AFQT score and having a high parental SES (or high parental income), the better choice would be to select having a high AFQT score. The fuzzy-set results underscore the fact that the choice is really about combinations of conditions—about recipes—not about individual variables. In short, choosing to not have a low AFQT score, by itself, does not offer protection from poverty. The configurational analysis presented in this chapter shows clearly that it is combined with other resources when it is linked to staying out of poverty.

Practical Appendix: Calibrations Used in the Fuzzy-Set Analysis

As previously noted, the calibration of fuzzy sets is central to fuzzy-set analysis. Miscalibrations distort the results of set-theoretic assessments. The main principles guiding calibration are that (1) the target set must be carefully defined and labeled and (2) the fuzzy set scores must reflect external standards based on both substantive knowledge and the existing research literature. While some might consider the influence of calibration decisions "undue" and portray this aspect of fuzzy-set analysis as a liability, in fact it is a strength. Because calibration is important, researchers must pay careful attention to the definition and construction of their fuzzy sets, and they are forced to concede that substantive knowledge is, in essence, a prerequisite for analysis. The fuzzy sets in the analysis presented in this chapter are degree of membership in the outcome, the set of individuals avoiding poverty, and degree of membership in sets reflecting various background characteristics and conditions. These are discussed in more detail below.

Avoiding poverty. To construct the fuzzy set of individuals avoiding poverty, this analysis uses the official poverty threshold adjusted for household size as provided by the NLSY, the same measure used by both Herrnstein and Murray (1994) and Fischer et al. (1996). In their analyses, both Herrnstein and Murray and Fischer et al. use the poverty status variable as a binary dependent variable in logistic regression analyses. However, their dichotomous measure places families with incomes just barely above the poverty level in the same category as those families with incomes far above the poverty threshold, such as comfortably upper-middle-class families. The fuzzy set procedure avoids this problem and is based on the ratio of household income to the poverty level for that household. Using the direct method for calibrating fuzzy sets (described in chapter 5), the threshold for full membership in the set of households not in poverty (fuzzy score = 0.95) is a ratio of 3.0 (household income is three times the poverty level for that household), the crossover point (fuzzy score = 0.5) is a ratio of 2.0 (household income is double the poverty level), and the threshold for full exclusion from the set of households not in poverty (fuzzy score = 0.05) is a ratio of 1.0 (household income is the same as the poverty level).

High school and college education. To measure educational attainment, the NLSY uses "Highest Grade Completed" (Center for Human Resource Research 1999, 138). This variable translates years of education directly into degrees (i.e., completing twelve years of education indicates a high school degree, while completing sixteen years completed indicates a college degree). Respondents with twelve or more years of school are fully in the set with a high school education (a fuzzy score of 1.0). On the other hand, those with only a primary school education (i.e., six years of school or less) are treated as fully out of the set of respondents with a high school education (a fuzzy score of 0.0). The fuzzy set thus embraces the six years of secondary school: 11 years = 0.75; 10 years = 0.60; 9 years = 0.45; 8 years = 0.30; and 7 years = 0.15. The fuzzy set of college-educated respondents was constructed by defining respondents with sixteen or more years of education as having full membership in the set with a college education (1.0), while those with twelve years of education or less were coded as fully out of the set (0.0). The in-between years were coded as follows: 13 years = 0.20; 14 years = 0.40; and 15 years = 0.60.

Parental income. The measure of parental income is based on the average of the reported 1978 and 1979 total net family income in 1990 dollars. It is the same measure used by Fischer et al. (1996) and was generously provided by Richard Arum. These data were used to create two fuzzy sets: the set of respondents with low parental income and the set of respondents with high parental income.

The fuzzy set of respondents with low parental income is similar in construction to the fuzzy set of households in poverty. First, the ratio of parents' household income to the poverty level was calculated using NLSY data on the official poverty threshold in 1979, adjusted for household size. Using the direct method of calibration described in chapter 5, it was determined that the threshold for full membership in the set with low parental income (0.95) is a ratio of 1.0 (parents' income is the same as the poverty level). Respondents with ratios less than 1.0 received fuzzy scores greater than 0.95. Conversely, the threshold for full exclusion (0.05) from the set with low parental income is a ratio of 3.0 (parents' household income was three times the poverty level). Respondents with ratios greater than 3.0 received fuzzy scores less than 0.05. The crossover point was determined to be two times the household-adjusted poverty level.

Multiples of the poverty ratio (household income divided by poverty level adjusted for household composition) were also used to construct the fuzzy set of respondents with high parental income. The threshold for exclusion from the set with high-income parents (0.05) is a ratio of three times the adjusted poverty level. The crossover point (0.50) was set at 5.5 times the adjusted poverty level, and the threshold for full membership was set at eight times the adjusted poverty level. The threshold for full membership corresponds roughly to three times the median family income, while the crossover point corresponds to roughly two times the median family income. Again, the direct method of fuzzy set calibration was used to calibrate degree of membership in this set.

Test scores. The AFQT scores used by Herrnstein and Murray (1994) are based on the *Armed Services Vocational Aptitude Battery*, which was introduced by the U.S. Department of Defense in 1976 to determine eligibility for enlistment. To construct the fuzzy-set measures of those with high AFQT scores and low AFQT scores, the analysis relies on categories used by the Department of Defense to place enlistees. Thus, the calibration of these fuzzy sets is grounded in practical decisions made by the military.

The military divides the AFQT scale into five categories based on percentiles. These five categories have substantive importance in that they determine eligibility for and assignment into different qualification groups. Persons in categories I (93rd to 99th percentile) and II (65th to 92nd percentile) are considered to be above average in trainability; those in category III (31st to 64th percentile) are about average; those in category IV (10th to 30th percentile) are designated as below average in trainability; and those in category V (1st to 9th percentile) are markedly below average. To determine eligibility for enlistment, the Department of Defense uses both aptitude and education as criteria. Regarding aptitude, the current legislated minimum standard is the 10th percentile, meaning that those who score in category V (1st to 9th percentile) are not eligible for military service. Furthermore, those scoring in category IV (10th to 30th percentile) are not eligible for enlistment unless they also have at least a high school education. Legislation further requires that no more than 20 percent of the enlistees be drawn from category IV, which further indicates that respondents in this category are substantially different from those in categories I to III.[5]

To construct the fuzzy set of respondents with low AFQT scores, respondents' AFQT percentile scores are used. The threshold for full membership (0.95) in the set of respondents with low AFQT scores was placed at the 10th percentile, in line with its usage by the military; respondents who scored lower than the 10th percentile received fuzzy

5. Of course, these standards are allowed to erode as the demand for military recruits increases.

membership scores greater than 0.95. The crossover point (0.5) was set at the 20th percentile, and the threshold for nonmembership was set at the 30th percentile, again reflecting the practical application of AFQT scores by the military. Respondents who scored better than the 30th percentile received fuzzy scores less than 0.05 in the set of respondents with low AFQT score.

The threshold for full membership (0.95) in the set of respondents with high AFQT scores was placed at the 93rd percentile, in line with the military's designation of the lower boundary of the highest category; the crossover point (0.5) was set at the 80th percentile; and the threshold for full nonmembership (0.05) in the set of respondents with high AFQT scores was placed at the 65th percentile, the bottom of the military's second highest AFQT category.

Household composition. Household composition has two main components: whether or not the respondent is married and whether or not there are children present in the household. All four combinations of married/not married and children/no children are present with substantial frequency in the NLSY data set. Respondents' marital status is coded as a crisp set, with a value of one assigned to those who were married in 1990. In general, married individuals are much less likely to be in poverty. While Fischer et al. (1996) use the actual number of respondents' children in 1990, having children is coded here as a crisp set. The rationale 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. The most favorable household composition, with respect to staying out of poverty, is the married/no children combination. The least favorable is the not-married/children combination.

REFERENCES

Achen, Christopher. 2005a. "Let's Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong." *Conflict Management and Peace Science* 22: 327–39.

———. 2005b. "Two Cheers for Charles Ragin.." *Studies in Comparative International Development* 40: 27–32.

Allison, Paul D. 1977. "Testing for Interaction in Multiple Regression." *American Journal of Sociology* 82: 144–53.

Amenta, Edwin, and Jane Duss Poulsen. 1996. "Social Politics in Context: The Institutional Politics Theory and Social Spending at the End of the New Deal." *Social Forces* 75: 33–60.

Becker, Howard S. 1958. "Problems of Inference and Proof in Participant Observation." *American Sociological Review* 23: 652–60.

Berg-Schlosser, Dirk. 2002. "Macro-Quantitative vs. Macro-Qualitative Methods in the Social Sciences—Testing Empirical Theories of Democracy." COMPASSS Working Paper 2002-2. www.COMPASSS.org/wp.htm.

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

Boswell, Terry, and Cliff Brown. 1999. "The Scope of General Theory. Methods for Linking Deductive and Inductive Comparative History." *Sociological Methods and Research* 28: 154–85.

Brady, Henry E. 2003. "Models of Causal Inference: Going beyond the Neyman-Rubin-Holland Theory." Paper presented at the Annual Meeting of Midwest Political Science Association, Chicago, April 4.

Brady, Henry, and David Collier, eds. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards.* Lanham, MD: Rowman and Littlefield.

Braumoeller, Bear. 2003. "Causal Complexity and the Study of Politics." *Political Analysis* 11: 208–33.

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