Research Note: Stepwise Consistency Analysis

Charles Ragin (cragin@uci.edu)

Background

Consistency scores are fundamental to the practice of Qualitative Comparative Analysis (Ragin 2006). They are used most often to gauge the degree to which a condition or combination of conditions constitutes a subset of an outcome. If a rough subset relation can be established, the researcher may interpret this connection as suggestive of a sufficiency relationship between the condition or combination of conditions and the outcome. In other words, the subset relation indicates that the cases sharing a specified condition or combination of conditions also share an outcome (Ragin 2008). For example, a researcher might assess whether having a strong left party is sufficient for the formation of a generous welfare state. If countries with strong left parties consistently established generous welfare states, the evidence is supportive of a sufficiency relationship.

With crisp sets, the calculation of subset consistency is straightforward. In the example just presented, it's simply the number of cases combining strong left parties and generous welfare states divided by the number of cases with strong left parties. With fuzzy sets (Ragin 2000; 2008), the calculation is:

$$\text{consistency of } X \leq Y = \text{sum}(\min(X_i,Y_i)) / \text{sum}(X_i)$$

where X_i is degree of membership in the condition or combination of conditions, and Y_i is degree of membership in the outcome. The numerator assesses the intersection of sets X and Y; the denominator allows expression of the intersection relative to the sum of the memberships in set X. The resulting consistency score indicates the proportion of set X that is contained within set Y. Consistency scores close to 1.0 provide clear evidence of a subset relationship, while scores less than 0.70 indicate substantial inconsistency. In general, if there are many cases with high membership in X coupled with low membership in Y, subset consistency scores are diminished accordingly.

Subset consistency scores are central to truth table analysis. The standard practice is to calculate the degree to which each combination of conditions, as specified by truth table rows, is a consistent subset of the outcome. Most applications of QCA use a consistency *threshold* to identify combinations of conditions with strong ties to the outcome. For example, a researcher might code truth table rows with consistency scores of 0.80 or greater as having strong ties to the outcome. Such rows would be coded "true" (1) on the outcome; rows with less than 0.80 consistency would be coded "false" (0).[1]

To illustrate, consider a truth table using data from the National Longitudinal Survey of Youth (NLSY79). As shown in table 1, the three conditions are race (white = 1; black = 0), gender (male = 1; female = 0), and class background (advantaged versus not-advantaged). Race and gender are conventional binary sets. Class background is operationalized as a fuzzy set based on the intersection of two other fuzzy sets, *educated-parent* and *not-low-parental-income*. To receive a high membership score in *advantaged-class-background*, a respondent's parent must be both educated and not low-income. The outcome is also a fuzzy set, degree of membership

---

[1] In practice, the coding of row outcomes as true or false comes after the deletion of truth table rows that do not meet whatever row frequency threshold the researcher has specified.

in the set of respondents *avoiding-poverty*. I describe the construction and calibration of the fuzzy sets in an appendix (see also Ragin and Fiss 2017).

Table 1: Impact of Race, Gender, and Advantaged Class Background on Poverty Avoidance

Row	Race: white = 1 black = 0	Gender: male = 1 female = 0	Class: Membership in Advantaged (1) or ~Advantaged (0)	Number of respondents	Consistency of poverty avoidance
1	1	1	1	1045	0.884
2	1	0	1	986	0.857
3	0	1	1	235	0.794
4	1	1	0	315	0.786
5	1	0	0	327	0.750
6	0	0	1	212	0.745
7	0	1	0	497	0.587
8	0	0	0	562	0.421

Three conditions yield a truth table with eight rows ($2^3 = 8$). The degree of membership of each case in each row is determined by the minimum of its memberships in the sets that make up the row. For example, a black female with a 0.3 membership in 'advantaged' has a membership of 0.3 in the row that combines black, female, and advantaged. This same case has a membership of 0.7 in the row that combines black, female, and not-advantaged (membership in not-advantaged = 1 – membership in advantaged), and it has a membership of 0 in the other six rows because black females have 0 membership in white and also 0 membership in male. Table 1 also reports the number of respondents in each row, with advantaged class background dichotomized at 0.5 (the cross-over point for fuzzy sets) to simplify the presentation.

Consistency of poverty avoidance is a set-analytic measure gauging the degree to which respondents with the combination of attributes in each row constitute a subset of respondents avoiding poverty (see Ragin 2008: chapter 3). In other words, it is an assessment of the degree to which the respondents in each row share the outcome in question—avoiding poverty. The consistency of poverty avoidance scores shown in the last column of table 1 are sorted from high to low, and range from 0.884 for *white•male•advantaged* to 0.421 for *black•female•~advantaged*.[2] As mentioned previously, it is standard practice for QCA users to establish a consistency threshold in order to distinguish combinations of conditions that are strongly linked to an outcome versus those that are not. Inspection of table 1 reveals a substantial gap in consistency scores between rows 6 and 7. Thus, one plausible consistency threshold for coding the outcome "true" (1) is 0.745. Using this threshold, the first six rows of table 1 are coded 1 on the outcome, while the bottom two are coded 0. Logical minimization of the coded truth table yields the following result:

avoidance of poverty ≥ white + advantaged

[2] Note: "~" indicates negation ("not-"); "•" indicates combined conditions (set intersection—logical and); "+" indicates alternate conditions or alternate combinations of conditions (set union—logical or).

This truth table solution states that respondents who are white or from advantaged class background are able to avoid poverty with at least moderate consistency (i.e., with scores of 0.745 or better). This set-analytic statement can be "clarified" (Ragin 2023:59-61) and rewritten as:

avoidance of poverty ≥ white + black•advantaged

by assigning the logical overlap between the two sets to "*white*." This restatement of the results of the truth table analysis makes it clear that all whites are able to avoid poverty with a consistency of 0.745 or better, but only blacks from advantaged class backgrounds experience comparable levels of insulation from poverty.

Stepwise Consistency Analysis

This research note presents an alternate, "stepwise" approach to the analysis of truth table consistency scores. In simple terms, the stepwise approach examines the consequences of lowering the consistency threshold row-by row beginning with the row with the highest consistency score. Table 2 illustrates the procedure. The initial consistency threshold is 0.884, row 1's consistency score (*white•male•advantaged*), which is the only row that is coded 1 on the outcome on the first step. The second threshold is pegged to row 2's consistency score, which lowers the consistency threshold to 0.857 and codes the first two rows to 1 on the outcome. The third threshold is pegged to row 3's consistency score, which lowers the threshold to 0.794 and codes the top three rows to 1 on the outcome, and so on through row 6, the last row with an acceptable consistency threshold. The last column of table 2 reports set-theoretic statements summarizing the stepwise results. The first cell in this column summarizes the first row results. The second cell summarizes the results of the union of the first two rows. The third cell summarizes the union of the first three rows, and so on through the remainder of the truth table. For example, the summary statement reported for the second row, *white•advantaged*, is based on the union of *white•male•advantaged* and *white•female•advantaged*. By progressively including more rows in a top-down, stepwise fashion, the scope of the conditions linked to poverty avoidance is gradually and methodically expanded, based on the rows with acceptable consistency scores.

Table 2: Stepwise Truth Table Results

Step	Conditions	Consistency	Cumulative Results (minimized)
1	white•male•advantaged	0.884	white•male•advantaged
2	white•female•advantaged	0.857	white•advantaged
3	black•male•advantaged	0.794	white•advantaged + male•advantaged
4	white•male•~advantaged	0.786	white•advantaged + male•advantaged + white•male
5	white•female•~advantaged	0.750	white + male•advantaged
6	black•female•advantaged	0.745	white + advantaged
7	black•male•~advantaged	0.587	white + advantaged + male
8	black•female•~advantaged	0.421	(all eight combinations are coded 1)

Stepwise consistency analysis provides a basis for interpreting the entire truth table, thereby avoiding the limitations that accompany focusing on a single outcome coding. For example, the first step confirms that the intersection of three favorable conditions (gender = male, race = white, and class = advantaged) offers the best insulation from poverty (see table 2). The second step indicates that for advantaged whites, gender is not a major factor when it comes to avoiding poverty. The third step extends the second step to embrace advantaged black males, yielding *male•advantaged* as a key combination of conditions. (The combination *black•female•advantaged* is not incorporated until step 6.) Step four offers the first appearance of not-advantaged as a condition, which yields the combination *white•male* as a recipe for avoiding poverty. It is noteworthy that white males are the first subpopulation to avoid poverty despite a lack of class advantage. Step 5 adds not-advantaged white females to the mix which in turn sets the stage for the debut of a single-ingredient recipe, *white*. Finally, step six incorporates the fourth combination that includes advantaged class background (*black•female•advantaged)*, which in turn permits the derivation of the second single-condition recipe, *advantaged*. Overall, the stepwise results (1) reinforce the importance of race and class, especially the benefits enjoyed by white males, and (2) highlight the disadvantages endured by black females.

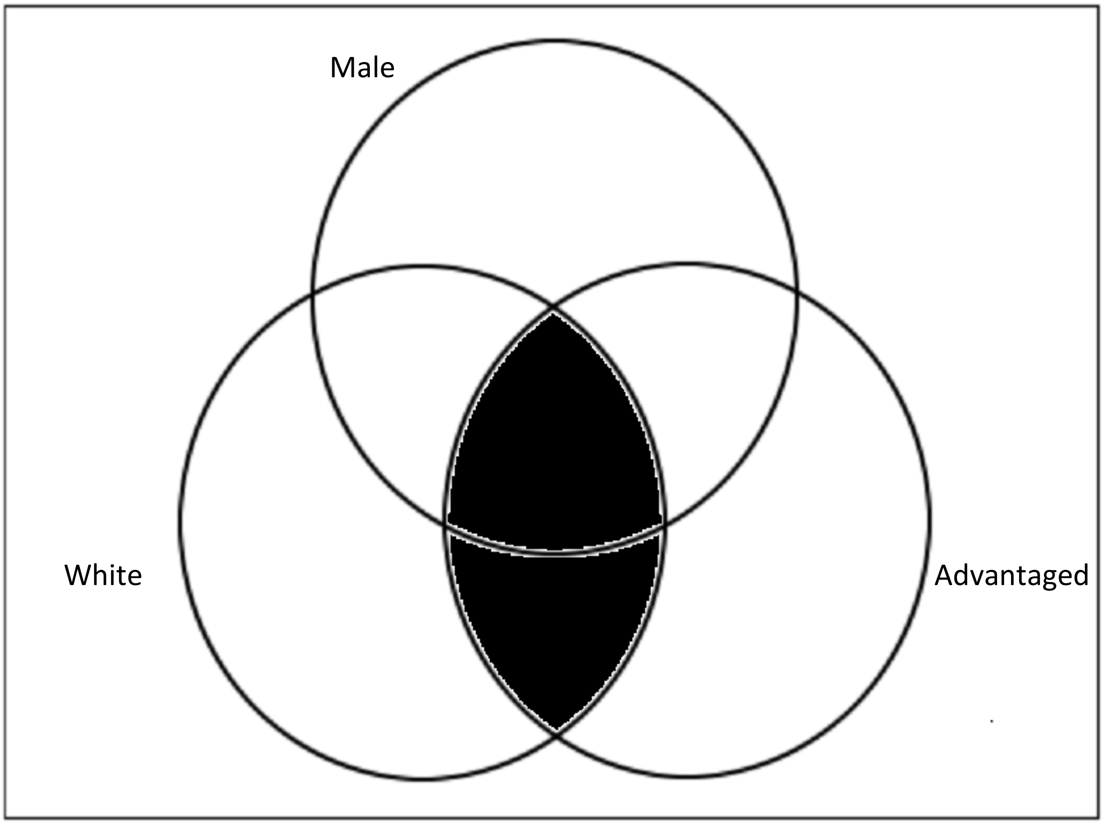
It is important to recognize that the truth table solutions listed in the last column of table 2 exhibit subset/superset relationships. The step 1 solution is a subset of the step 2 solution, which in turn is a subset of the step 3 solution, and so on. The subset relations are most visible when represented using Venn diagrams. Compare figures 1, 2 and 3. Figure 1 models step 2's solution—the intersection of white and advantaged. Figure 2 models step 4's solution and illustrates the three two-way set intersections that are linked to the avoidance of poverty. Figure 3 models step 6's solution—the union of white and advantaged.

This nuanced analysis of consistency scores is especially appropriate in situations where there is no obvious threshold value. For example, sometimes there is an array of consistency scores in the 0.75 to 0.90 range, offering several plausible threshold values. Rather than focusing on a single, possibly arbitrary threshold value, it is preferable to conduct a stepwise examination of the truth table, using multiple, rank-ordered consistency thresholds. Stepwise analysis also protects researchers from the charge that they have selected a threshold value in an opportunistic manner, in order to favor a specific interpretation of the evidence.

Limitations

Many, if not most, truth tables are straightforward and can be deciphered adequately using a single threshold value. Some truth tables, for example, have only a single row with a high consistency score. Stepwise consistency analysis is useful only when there are at least two rows with high consistency scores. Furthermore, the stepwise approach is likely to be less useful when Ns are small or when most truth table rows have few or no cases. The ideal context for the stepwise procedure is (1) a moderate-to-large number of cases, (2) a relatively well-populated truth table, and (3) several plausible consistency threshold values.

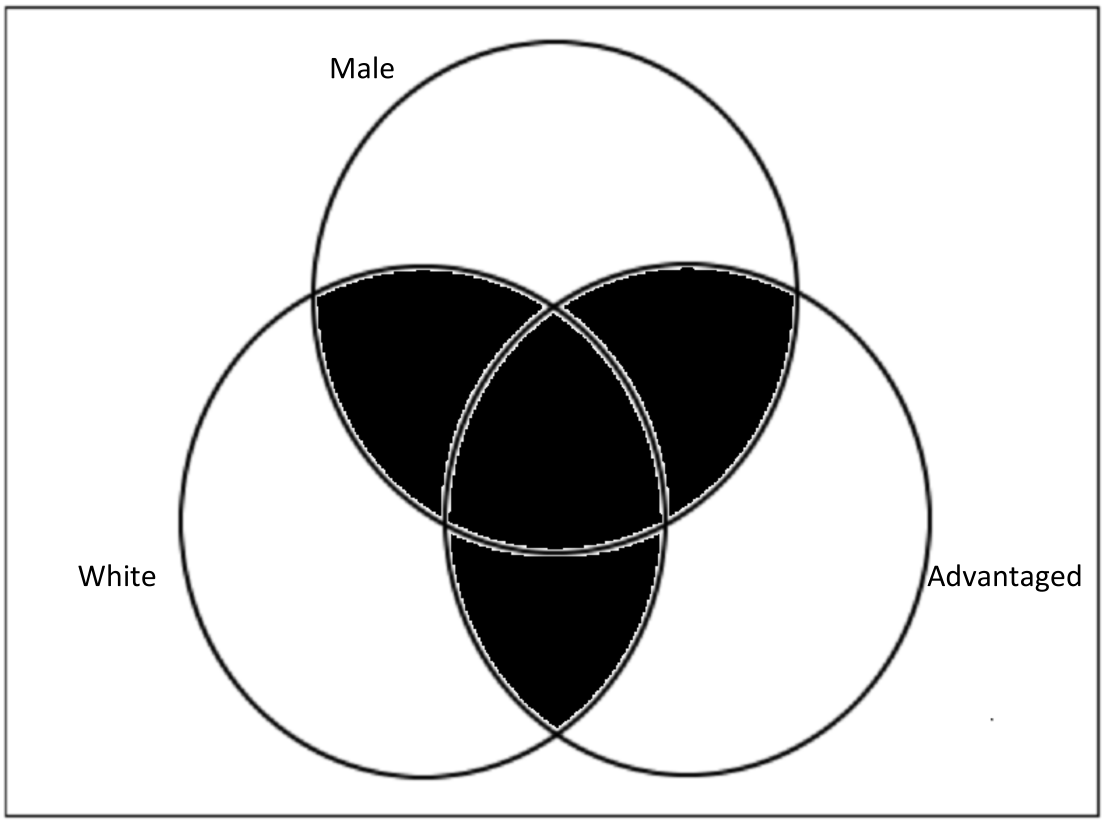
Figure 1: Venn Diagram for Truth Table Solution at Step 2.



avoidance of poverty ≥ white•advantaged

consistency = 0.857

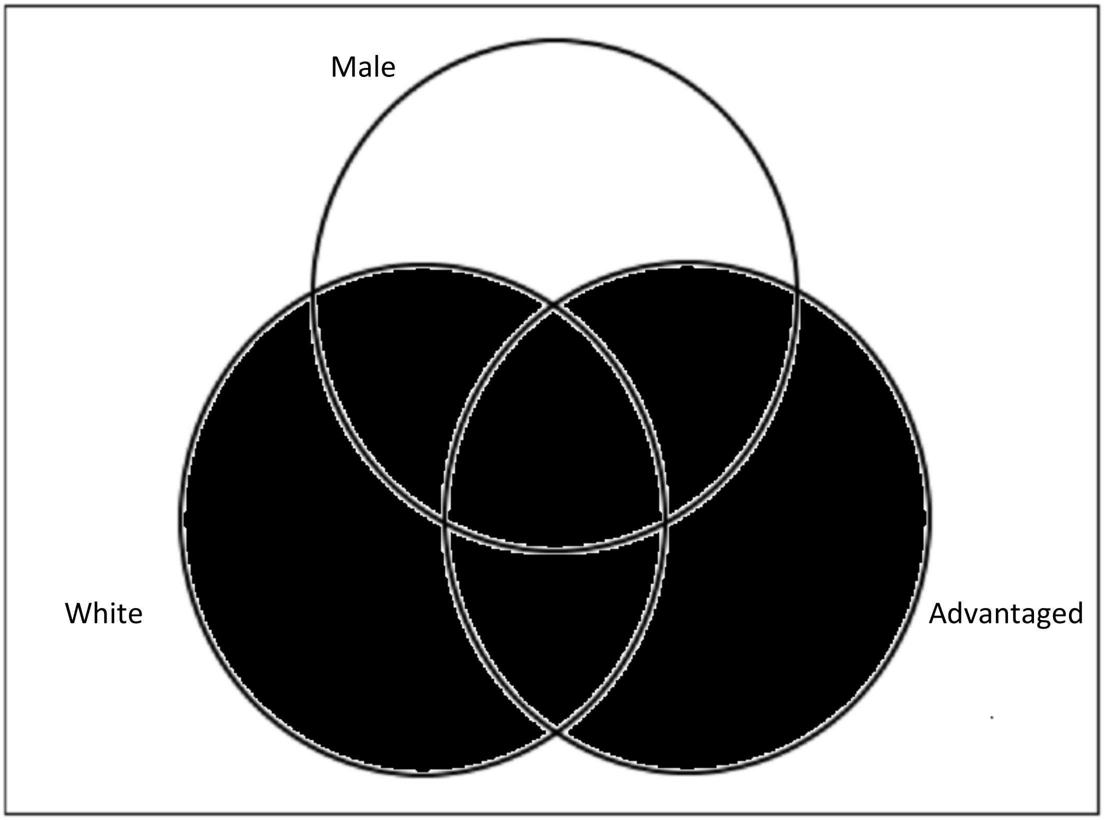
Figure 2: Venn Diagram for Truth Table Solution at Step 4.



avoidance of poverty ≥ white•advantaged + male•advantaged + white•male

consistency = 0.786

Figure 3: Venn Diagram for Truth Table Solution at Step 6.



avoidance of poverty ≥ white + advantaged

consistency = 0.745

Appendix: Calculating Fuzzy Set Membership Scores

Avoiding Poverty. To construct the fuzzy set of individuals avoiding poverty, I first calibrate respondents' degree of membership in poverty. I base the measurement on the official poverty threshold adjusted for household size and composition. The official poverty threshold is an absolute threshold, meaning it was fixed at one point in time and is updated solely for price changes.

The measure of poverty is based on the ratio of household income to the poverty level for that household. Using the direct method for calibrating fuzzy sets (see Ragin 2008, chapter 5), the threshold for full membership in the set of households *in-poverty* (fuzzy membership score = 0.95) is a ratio of 1.0 (household income is the same as the poverty level); the cross-over point (fuzzy membership 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 *in-poverty* (fuzzy membership score = 0.05) is a ratio of 3.0 (household income is three times the poverty level for that household).

The fuzzy set of households in poverty is a symmetric set; that is, it is truncated at both ends and the crossover point is set exactly at the halfway mark between the thresholds for membership and non-membership. Thus, the set of respondents *avoiding-poverty* is based on a straightforward negation of the set of respondents *in-poverty*. With fuzzy sets, negation is accomplished simply by subtracting membership scores from 1.0 (see Ragin 2008, chapter 2). That is, (membership in *avoiding-poverty*) = 1 − (membership *in-poverty*). For example, a case that is mostly but not fully *in* the set of respondents *in-poverty*, with a score of 0.90, is mostly but not fully *out* of the set of respondents *avoiding-poverty*, with a score of 0.10.

The use of a ratio of three times the poverty level for full membership in the set of cases avoiding 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 Administration 1998, table 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, table F-7).

Advantaged Class Background. To assess degree of membership in *advantaged-class-background*, I combine two fuzzy sets, *not-low-income parents* and *educated-parents*. To obtain greater than 0.5 membership (i.e., more in than out) in *advantaged-class-background*, respondents must have greater than 0.5 membership in both *not-low-income-parents* and *educated-parents*. I describe the calibration of *not-low-income-parents* first.

I assess parental income by first 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. The fuzzy set of respondents with not-low-income parents is similar in its construction to the fuzzy set of households avoiding poverty, described previously. That is, I first calculate the ratio of parents' household income to the poverty level, using NLSY data on the official poverty threshold in 1979, adjusted for household size and composition. Using the direct method of calibration, the threshold for full membership in the set with not-low parental income (.95) is a ratio of 5.5 (parents' income was five and a half times the poverty level). Respondents with ratios greater than 5.5 times receive fuzzy scores between 0.95 and 1.0. Conversely, the threshold for full exclusion (.05) from the set with not-low parental income is a ratio of 2 (parents' household income was only double the poverty level). Respondents with ratios less than 2 received fuzzy scores between 0.05 and 0. The cross-over point (membership = 0.50) is pegged at three times the household-adjusted poverty level.

I use the greater of mother's and father's years of education to assess parental education. In order to convert this variable to a fuzzy set, I use the following benchmarks: Parents with 12 or more years of schooling are more in than out the set of *educated-parents* (fuzzy score > 0.5). Those with fewer than 9 years of education are treated as fully out of the set (fuzzy score of 0), and those with 16 or more years of education are treated as fully in the set. The specific translation scheme is 0-8 years = 0; 9 years = 0.1; 10 years = 0.2; 11 years = 0.4; 12 years = 0.6; 13 years = 0.7; 14 years = 0.8; 15 years = 0.9; 16+ years = 1.0.

These two fuzzy sets, *not-low-income-parents* and *educated-parents* are joined by logical *and* (set intersection) to create *advantaged-class-background*. To implement logical *and* with fuzzy sets, it is necessary simply to take the lower of the two membership scores.

References

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

Ragin, Charles C. 2000. *Fuzzy Set Social Science*. Chicago: University of Chicago Press.

Ragin, Charles C. 2006. "Set Relations in Social Research: Evaluating Their Consistency and Coverage." *Political Analysis* 14(3):291-310.

Ragin, Charles C. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Berkeley, CA: University of California Press.

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

Ragin, Charles C. 2023. *Analytic Induction for Social Research*. Berkeley, CA: University of California Press.