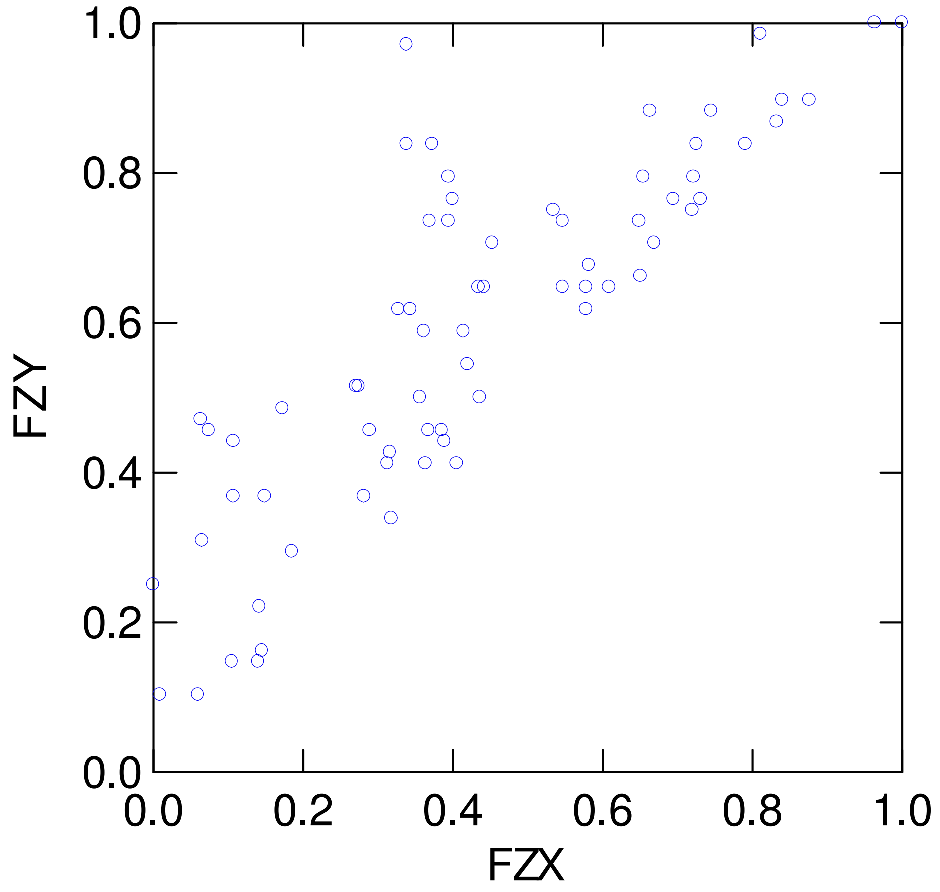
PRI Consistency (“PRE” in fsQCA’s Truth Table Spreadsheet)

PRE stands for “proportional reduction in error.” Many measures of association (e.g., both lambda and r^2) are PRE measures. PRI stands for “proportional reduction in inconsistency” and is an alternate measure of the consistency of subset relations in social research. It is more exacting than the usual measure (“raw” consistency) and is applied only to fuzzy sets.

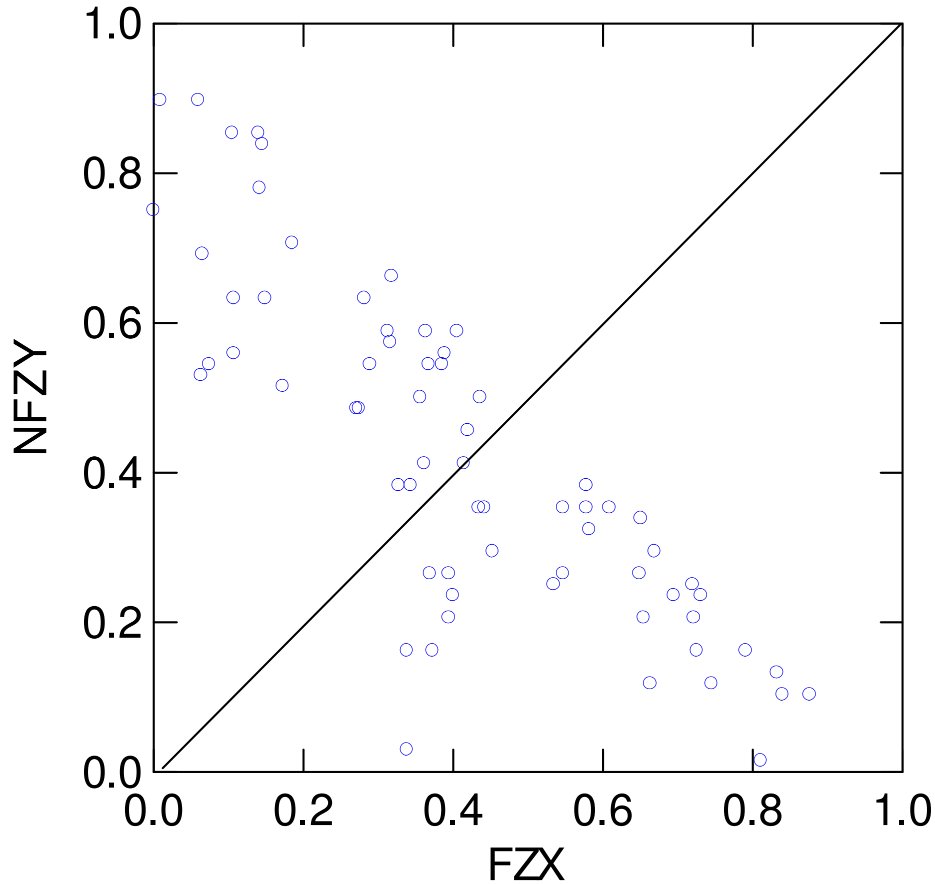
With fuzzy sets (but not crisp) it is possible for a causal condition to be a subset of the outcome and a subset of the negation of the outcome. Consider the following membership scores:

Causal Condition **x**	**Outcome** **y**	**~Outcome** **~y**
.4	.4	.6
.4	.5	.5
.3	.6	.4
.3	.7	.3
.2	.3	.7
.2	.4	.6
.1	.8	.2

Notice that $x \leq y$ is true; so is $x \leq \sim y$, indicating that x is a subset of y and a subset of ~y. Notice also that x is always ≤ .5. If x is greater than .5, then it is impossible for x to be less than y and less than ~y. These fuzzy subset relations can also be represented graphically. Here’s the plot for $x \leq y$:

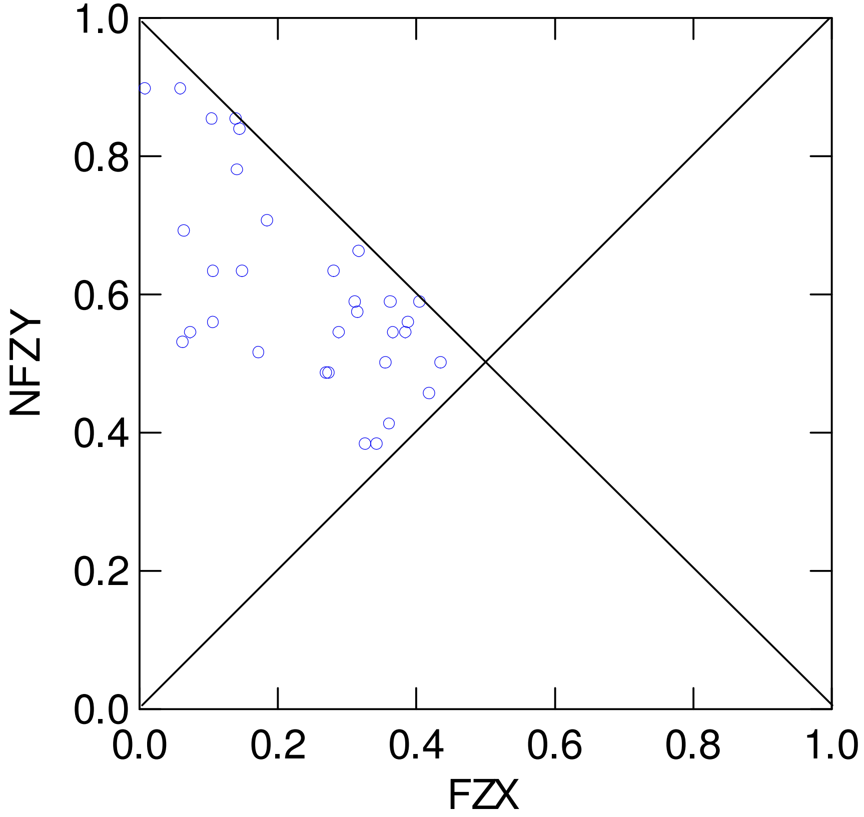


This shows the characteristic subset relations. Next, using the same two fuzzy sets, let negate y and redo the plot:



You can see that many of the points are still above the main diagonal, indicating that some points are consistent with both subset relations. Notice also that these points are all $x \leq .5$.

It is possible to divide the scatterplot into four sectors and identify the sector of the plot where : x ≤ y and : x ≤ ~y:



Any case that fall in the left triangle satisfy both x ≤ y and x ≤ ~y. The key triangle from the viewpoint of the consistency of x ≤ y is the top triangle. Cases in this triangle strongly violate x ≤ ~y, while strongly supporting x ≤ y. The reverse is true for the bottom triangle. Cases in the right most triangle violate both x ≤ y and x ≤ ~y.

From a computational point of view, the key is to eliminate the influence of cases that satisfy both $x \leq y$ and $x \leq \sim y$ from the computation of the consistency of the evidence with the fuzzy subset relation. At a minimum, it is good to know both the “raw consistency” score and the consistency of the evidence when disregarding the cases that satisfy both $x \leq y$ and $x \leq \sim y$.

Recall that raw consistency is computed as follows:

$$\frac{\sum \min(x_i,y_i}{\sum(x_i)}$$

It indicates the degree to which the x membership scores are consistently less than the y membership scores. One simple way to remove the influence of cases that satisfy both $x \leq y$ and $x \leq \sim y$ is to subtract $\sum \min(x_i,y_i,\sim y_i)$ from both the numerator and the denominator. This sum is the fuzzy intersection of sets x, y, and ~y.

$$\frac{\sum \min(x_i,y_i) - \sum \min(x_i,y_i,\sim y_i)}{[\sum(x_i) - \sum \min(x_i,y_i,\sim y_i)]}$$

Essentially, this calculation removes the three-way intersection of x, y, and ~y from the computation of consistency.

Proportional Reduction Interpretation

The $\sum x$ has two parts: the consistent part and the inconsistent part (with respect to it subset relation with y).

The consistent part: $\sum \min(x_i, y_i)$
The inconsistent part: $\sum (x_i) - \sum \min(x_i, y_i)$

The same is true with respect to the subset relation between x and the intersection of y and ~y, that is, there is part of x that is consistent and a part that is consistent:

The consistent part: $\sum \min(x_i, y_i, \sim y_i)$
The inconsistent part: $\sum (x_i) - \sum \min(x_i, y_i, \sim y_i)$

Now the fun part: Proportion reduction errors have a characteristic structure, namely:

$$\frac{(E_1 - E_2)}{E_1}$$

Where E_1 is the sum of the errors you make when you lack some key knowledge, and E_2 is the sum of the errors you make once you have this knowledge. Ideally, this knowledge helps you make fewer errors. The usual PRE situation is that when calculating E_1 you have knowledge only of the distribution of the dependent variable; E_2 adds knowledge of the cases' scores on one or more independent variables.

For set theoretic analysis, we treat E_1 as the situation where you don't know if x is a subset of y or ~y, and E_2 is the situation where you test the idea that x is a subset of y.

We treat the inconsistent parts of x sketched above as our calculations of E_1 and E_2:

$$E_1 = \sum(x_i) - \sum \min(x_i, y_i, \sim y_i)$$
$$E_2 = \sum(x_i) - \sum \min(x_i, y_i)$$

In PRE format, the computation becomes:

$$\frac{[\sum(x_i) - \sum \min(x_i, y_i, \sim y_i)] - [\sum(x_i) - \sum \min(x_i, y_i)]}{\sum(x_i) - \sum \min(x_i, y_i, \sim y_i)}$$

which simplifies to:

$$\frac{[\sum \min(x_i, y_i) - \sum \min(x_i, y_i, \sim y_i)]}{\sum(x_i) - \sum \min(x_i, y_i, \sim y_i)}$$

which is the measure of PRI consistency shown previously.