PITFALLS IN QCA'S CONSISTENCY MEASURE

Tim HAESEBROUCK[1]

Over the years, Qualitative Comparative Analysis developed into a widely-used analytical technique in political science. This article, however, reveals that the consistency measure, QCA's single most important parameter of fit, is significantly flawed. Contrary to the requirements that were set forth when this measure was introduced, inconsistent cases with small membership scores exert greater bearing on the consistency score than inconsistent cases with large membership scores. In consequence, the measure does not accurately express the degree to which empirical evidence supports statements of sufficiency and necessity. After revealing this flaw, the article introduces a new formula for calculating consistency, which more accurately assesses the evidence for sufficiency and/or necessity. Subsequently, it demonstrates how the standard consistency measures leads to the misinterpretation of empirical evidence by reanalysing two recent QCA-application.

Key words: Research Methods, QCA, Consistency, Fuzzy Sets.

1 INTRODUCTION

In the almost thirty years since the publication of Charles Ragin's "The Comparative Method", Qualitative Comparative Analysis (QCA) has developed into a widely-used analytical technique in political science. The number of QCA-related articles published in peer-reviewed journals is increasing exponentially, from forty-five in 2012 to no fewer than ninety-nine in 2013 (Marx, Rihoux and Ragin 2014, 115; Rihoux 2014). Over the years, the technique went through numerous modifications and adjustments. One of the most important developments was the introduction of the consistency measure, which eventually became QCA's single most important parameter for assessing sufficiency and necessity (Wagemann and Schneider 2010, 289). Strikingly however, this formula does not meet the requirements Ragin (2006) formulated when he introduced the measure in its current form. Contrary to the latter's assertions, small disconfirming cases have greater bearing on the consistency score than large disconfirming cases. Consequentially, the standard consistency

[1] **Tim HAESEBROUCK**, Master in EU-studies (Ghent University 2010), is a PHD researcher at Ghent University, Belgium. His research interests include Qualitative Comparative Analysis, the Responsibility to Protect and the EU's Common Security and Defence Policy. He has previously published in International Politics and European Foreign Affairs Review. Contact: tim.haesebrouck@ugent.be.

measure does not adequately express the degree to which the empirical data is in line with statements of sufficiency and/or necessity.

After revealing this flaw in QCA's most important parameter, this article demonstrates how it leads to the misinterpretation of empirical evidence for sufficiency and necessity and introduces a formula that more accurately assesses the evidence for sufficiency and/or necessity. The article is structured around three main parts. First, the general purpose of calculating consistency is described. Subsequently, I demonstrate that the standard formula does not meet all requirements Ragin deemed necessary to achieve this purpose and introduce a new formula, which more accurately assesses the evidence for sufficiency and/or necessity. Finally, two recent applications of fuzzy set QCA, Mello (2014) and Schneider and Makszin (2014), are used to illustrate the impact of the flaw on empirical research and the benefits of using the alternative consistency measure.

2 THE CONSISTENCY MEASURE

QCA is generally used to establish set-theoretic connections between one case property, defined as the outcome, and other properties, defined as the causal conditions (Wagemann and Schneider 2010, 380). As extensively demonstrated in Ragin (2000, 203–260; 2008, 13–28) and Schneider and Wagemann (2012, 56–91), such subset relations are intimately linked to the notions of sufficiency and necessity. Since a condition is sufficient if the outcome is always present when this condition is present, the set defined by a sufficient condition constitutes a subset of the set defined by the outcome. Inversely, a condition is necessary if it is always present when the outcome is present. Therefore, the set defined by a necessary condition constitutes a superset of the set defined by the outcome.

The assessment of set-theoretic connections is straightforward in the original crisp set version of QCA. Cases are either present or absent in a crisp set, respectively indicated by a value of 1 and 0. In consequence, establishing a set-relation solely requires examining whether each case with a score of 1 in the alleged subset also has a score of 1 on the outcome. This straightforward procedure cannot be duplicated in the more sophisticated fuzzy set QCA, in which membership scores can vary between full membership (value of 1) and full non-membership (value of 0). In fuzzy sets, assessing subset relations requires examining whether each case's membership score in subset X is consistently equal or less than its corresponding score in superset Y, thus whether $X \leq Y$.

Perfect subset relations and fully necessary or sufficient conditions are relatively rare in social science (Ragin 2000, 108). This inspired Ragin to introduce the consistency-parameter, which provides a descriptive measure of the degree a perfect set relation is approximated (Ragin 2006, 292; Wagemann and Schneider 2010, 389). It is predominately, but not exclusively, used in fuzzy set QCA, and therefore designed to assess the extent to which $X \leq Y$ (Ragin 2008, 39).

The first consistency measure was introduced in Ragin's (2000) "Fuzzy Set Social Science". The original formula was very straightforward, it simply calculated the proportion of the cases where $X \leq Y$:

$$\frac{N(X \leq Y)}{N}$$

If X ≤ Y in all cases, this formula yields a score of 1. The higher the proportion of inconsistent cases, the closer it approaches 0.

In subsequent publications, Ragin (2006) adjusted the formula twice. First, he asserted that cases with strong membership scores in the subset are more relevant than cases with weak membership scores:

"a case with a membership of only 0.25 in the set of cases with the causal combination (X) and a score of 0.0 in the outcome set (Y) is just as inconsistent as a case with a score of 1.0 in the causal combination an a score of 0.75 in the outcome. In fact however, the second inconsistent case, with full membership in X, clearly has more bearing on the set-theoretic argument because it is a much better instance of the causal combination. It thus constitutes a more glaring inconsistency than the first case…" (Ragin 2006, 295).

In order to take into account the size of the membership scores, Ragin introduced a new formula:

$$\frac{\sum \text{Membership Scores Consistent Cases in X}}{\sum \text{Membership Scores All Cases in X}}$$

This formula was further refined to meet a second requirement: near misses should have less bearing on the result than membership scores that "exceed their target by a wide margin" (Ragin 2006). Since larger inconsistencies more strongly contradict the existence of a subset relations, the size of the inconsistent portions should have an impact on the consistency measure. In order to take this into account, Ragin added the consistent parts of inconsistent cases to the numerator. This was formalized in the following formula:

$$\text{Consistency}(X_i \leq Y_i) = \sum(\min(X_i, Y_i) / \sum(X_i)$$

The consistency measure was thus developed to evaluate the empirical support for set-theoretic relationships, and thus sufficiency and/or necessity. As argued by Ragin when introducing the different formulas for calculating consistency, this depends on (1) the ratio between consistent and inconsistent cases, (2) the relative size of these consistent and inconsistent cases and (3) the size of the inconsistencies. In order to meet these requirements, the consistency measure was adjusted twice. The second adjusted formula is currently the standard measure for consistency. It is presented in all major text books on QCA (Ragin 2008, 45–54; Rihoux and Ragin 2009, 102; Schneider and Wagemann 2012, 123–129 and 139–144), used in nearly all fsQCA-applications and incorporated in the popular fsQCA-software (Ragin and Davey 2012) as well as the more sophisticated QCA-package for R (Thiem and Dușa 2013).

3 FLAW IN THE STANDARD CONSISTENCY MEASURE

Strikingly however, the standard formula for calculating consistency is afflicted by a significant flaw. Although this formula was explicitly developed to meet the three criteria described above, it fails to fully meet the second criterion. Contrary to Ragin's assertions, cases with a larger membership score in the subset not always have greater bearing on the result of this formula. In fact, they only exert a greater impact when they confirm a set theoretic connection. Cases with stronger membership scores that disconfirm a set-relation however, ceteris paribus, have less impact on the consistency score.

This can be illustrated with the example Ragin (2006, 295) used to demonstrate the need for making the first adjustment to the formula for calculating consistency. Table 1 represents two datasets, in which only the membership score of case 4 diverges. In both datasets, case 4 contradicts X being a subset of Y to the same extent: X exceeds Y by 0.25. However, it has a stronger membership scores in X in dataset 1. According to Ragin, (see citation above), it therefore constitutes a more glaring inconsistency. The standard consistency formula however yields a higher value for X_1 as a subset of Y_1 than for X_2 as a subset of Y_2, respectively 0.9 (2.25/2.5) and 0.86 (1.5/1.75).

TABLE 1: SMALL VS. LARGE INCONSISTENT CASES

		Data set 1				Data set 2		
Case		X_1	Y_1	(min X,Y)		X_2	Y_2	(min X,Y)
1		1	1	1		1	1	1
2		0.25	0.25	0.25		0.25	0.25	0.25
3		0.25	0.25	0.25		0.25	0.25	0.25
4		1	0.75	0.75		0.25	0	0
Sum		2,5		2.25		1.75		1,5

The above example is not a carefully picked out exception. The membership scores Ragin uses to motivate the first adjustment of the measure are as good as any for current purpose, since smaller inconsistent cases always have greater bearing on the consistency score. This is a consequence of the fact that cases with larger membership scores in the subset, ceteris paribus, have a larger consistent part. Therefore, they add relatively more to the numerator than the denominator of the consistency formula.

The currently used formula for consistency thus does not fully meet Ragin's own requirements. To keep this flaw from inducing wrong conclusions on necessity and/or sufficiency, it seems advisable not to rely on the standard formula. Researchers could calculate consistency with the first adjusted formula. Cases with a stronger membership score in the subset always have a greater bearing on the result of this formula. This can be illustrated using the datasets from the above example. Both datasets contain the same three consistent cases, which sum equals 1.5. The sum of all the membership scores in X amounts to 2.5 in "data set 1", while it only equals 1.75 in "data set 2". Dividing the membership scores of the consistent cases by all cases therefore results in a much lower value for X_1 as a subset of Y_1 than for X_2 as a subset of Y_2, respectively 0.6 (1.5/2.5) and 0.86 (1.5/1.75). Unfortunately, this formula prescribes the same penalty for large and small inconsistencies. In consequence, small inconsistencies can have a disproportionately large impact on the consistency score. Case 4, for example, has a very large impact on the consistency of X_1 as a subset of Y_1, although its membership score in X_1 exceeds its score in Y_1 only by a relatively narrow margin.

A more optimal solution is to adjust the standard formula to increase the impact of the inconsistent portion of the cases with a high membership score in X. In the standard formula, the numerator equals the sum of the consistent portions of the cases, which is formalized as min (X, Y). The denominator of the standard formula consists of the sum of the membership scores of the cases in X, which equals the sum of the consistent and the inconsistent portions. The standard formula can thus be presented as follows:

$$\frac{\sum(\text{Consistent portion } X_i)}{\sum(\text{Consistent portion } X_i + \text{ Inconsistent portion } X_i)}$$

To increase the impact of the inconsistent portion of the cases with a high membership score in X, these are multiplied by the corresponding membership score in X. Evidently, this product will be higher for cases with a large membership score in X. As a result, a case with a strong membership score in X will have greater bearing on the consistency score. However, since membership scores can never exceed 1, this product will generally be smaller than the inconsistent portion of X. In consequence, inconsistencies will generally have less impact on the consistency score. This can be avoided by taken the square root of the product:

$$\sqrt{(\text{Inconsistent portion } X_i) * X_i}$$

The resulting formula can be formalized as follows:

$$\frac{\sum(\min(X_i, Y_i)}{\sum(\min(X_i, Y_i) + \sqrt{\max(X_i - Y_i, 0) * X_i})}$$

In line with the standard formula, the new consistency measure subscribes greater penalties for large inconsistencies: "max (X_i - Y_i, 0)" will add more to the denominator if X exceeds Y by a wider margin. However the inconsistent cases with a large membership score in X will have more impact on the consistency score, since the inconsistent portions are multiplied by X_i. This can be illustrated with the dataset from the above example. Both datasets have one inconsistent case with an inconsistent portion of 0.25. The membership score of this case in X equals 1 in dataset 1 and 0.25 in dataset 2. In consequence, the square root of the product of X and the inconsistent portion of X is higher in dataset 1 than in dataset 2, respectively 0,5 () and 0.25 (). The inconsistency of case 4 will thus add more to the denominator in dataset 1 than in dataset 2. In consequence, the new formula yields a lower value for X_1 as a subset of Y_1 than for X_2 as a subset of Y_2, respectively 0.82 (2.25/(2.25 + 0.5)) and 0.86 (1.5/1.5 + 0.25).

The new formula thus combines the strengths of both previous consistency measures. In line with the standard formula, it takes the size of the inconsistencies into account; in line with the first adjusted formula, it attributes greater impact to large inconsistent cases.

4 THE CONSISTENCY MEASURE AND APPLIED QCA

Two recently published studies serve as illustrations on how the identified shortcoming in the standard consistency measure affects the results of applications of fsQCA. The first is drawn from a book by Mello (2014), the second from an article by Schneider and Makszin (2014). Both of these were published very recently and (co-) written by scholars that can be considered experts in the field of QCA. To my best knowledge, they are flawless applications of QCA, that meet all current standards of good practice. This section thus certainly does not constitute a critique on these studies. Instead, it illustrates how the currently used consistency measure leads to misguided conclusions, even in the best applications of QCA.

Before considering the two examples, the general procedure followed in fsQCA must be briefly explained.[2] Rather than focusing on single necessary or sufficient conditions, QCA is used to establish a more complex form of causal relations, generally captured under the expression "multiple conjunctural causation". In line with the notion of conjunctural causation, QCA accounts for the possibility that single conditions are not sufficient to produce an outcome on their own, but are sufficient in combination with other conditions. In line with the idea of multiple causation or equifinality, QCA allows to take into account the possibility that several of such combinations are sufficient for the same outcome. The key tool for establishing such complex causal relations is the "truth table", which contains a row for every possible combination of conditions. At the first stage of the analytical procedure, each case's membership score in these rows is calculated with fuzzy multiplication. Rows without cases with a membership score above 0.5 correspond to logical remainders, combinations of conditions that lack good empirical instances. For every row that does not correspond to such a remainder, the researcher has to decide whether it can be considered sufficient for the outcome.

The assessment of sufficiency is based on the consistency of the row as a subset of the outcome. Rows are considered sufficient, and assigned a score of 1 in the outcome column, if their consistency exceeds a cut-off point (Schneider and Wagemann 2012, 279). After the rows have been assigned an outcome value, the truth table is reduced with Boolean minimization. Depending on the logical remainders that are incorporated in the minimization procedure, minimization can result in different types of solutions (Ragin 2008, 145–177). However, each of these depends on which truth table rows were considered sufficient for the outcome, which, in turn, depends on their consistency.

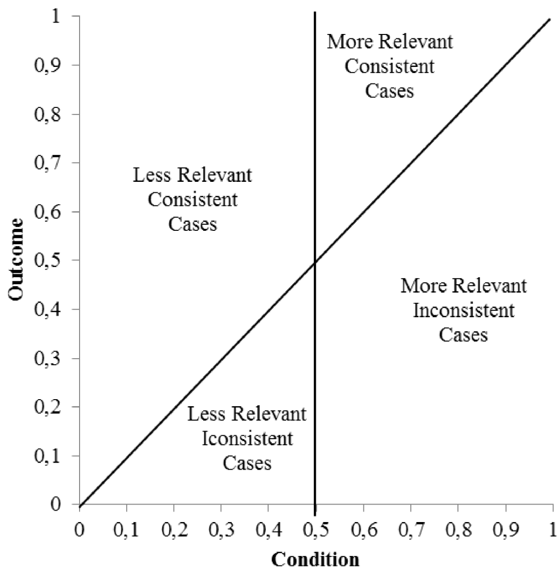
The flaw in the currently used consistency measure can however cause the consistency of truth table rows with relatively large inconsistent cases to exceed the consistency cut-off point, while rows with equally strong evidence for sufficiency, but smaller inconsistent cases, might fall below this threshold. In consequence, the larger impact of small inconsistent cases might cause researchers to code the former as sufficient for the outcome, while coding the latter as insufficient. The resulting formulas will hereby not be fully in line with the empirical evidence at hand.

In both examples, the coding of the truth table rows does not reflect the empirical evidence for sufficiency, which affects the validity of the resulting solutions. To evaluate whether the evidence for sufficiency is in line with the three requirements set out by Ragin when he developed the consistency formulas (cf. supra), the fuzzy membership scores in the truth table rows and the outcome are depicted in x-y plots (Schneider and Grofman 2006; Schneider and Rohlfing 2013). The scores in the outcome are displayed on the y-axis, the scores in the truth table row on the x-axis. The diagonal defines a line on which cases have equal membership scores in X and Y. Since the membership scores in X are smaller or equal to the corresponding scores in Y in all cases on and above this line, all cases in this area are consistent with the statement that X is a subset of Y. Similarly, X is larger than Y in all cases below this line, which are thus inconsistent with the statement of sufficiency. The degree to which these

[2] For an extensive treatment, see Schneider and Wagemann (2012, 91–116) and Ragin (2008, 124–144).

cases are inconsistent depends on the distance towards this diagonal. The further inconsistent cases are situated from the diagonal, the larger their inconsistent parts. The vertical line goes straight up from the 0.5 value of the x-axis and distinguishes relatively relevant from relatively irrelevant observations. Since cases with a membership above 0.5 are generally considered good instances of a causal condition, cases situated at the right of this axis can be considered to have relatively strong membership scores and, consequentially, constitute more relevant observations. The distinction between relevant-irrelevant and consistent-inconsistent cases that results from the intersection of the vertical and diagonal is graphically depicted in the x-y plot in figure 1.

FIGURE 1: X-Y PLOT TYPES OF CASES



4.1 Mello: Democratic participation in armed conflict

The first example is drawn from a book length study on democratic participation in armed conflict. The author, Patrick Mello, has published articles that apply fsQCA in journals included in the Thomson Reuters citation index, has written a commented review of QCA applications and teaches a course on set theoretic methods in the renowned ECPR summer school (Mello 2012, 2013). He can thus be considered an expert on the method. As could be expected, the QCA-application in this example is flawless, so the misinterpretation of the evidence can be fully attributed to the flaw in the consistency measure. The goal of the fsQCA was to determine the conditions under which democracies participated in "Operation Allied Force", NATO's 1999 military intervention in Kosovo. 23 cases were included in the analysis and compared on five explanatory conditions: military power (M), parliamentary veto (V), constitutional restrictions (C), public support (S) and executive ideology (E). The resulting truth table is presented on the left hand side of table 2. The consistency threshold was set at 0.87. Consequentially, rows 1-6 were considered sufficient for the outcome, rows 7-15 insufficient.

TABLE 2: TRUTH TABLE MELLO (2014)

	Conditions					Original Analysis		New Analysis		
row	M	V	C	S	E	Standard Consistency	MP	Previous Consistency	Adjusted Consistency	MP
1	1	0	0	1	1	1	1	1	1	1
2	1	1	0	0	1	1	1	1	1	1
3	0	0	1	1	0	1	1	1	1	1
4	0	0	0	1	0	0.99	1	0.93	0.98	1
5	0	0	0	0	1	0.93	1	0.84	0.89	1
6	0	0	0	1	1	0.87	1	0.81	0.84	1
7	0	1	0	1	1	0.83	0	0.64	0.77	0
8	1	1	1	1	0	0.8	0	0.8	0.8	1
9	0	0	0	0	0	0.77	0	0.63	0.73	0
10	0	1	0	0	1	0.66	0	0.5	0.62	0
11	0	1	0	0	0	0.66	0	0.49	0.61	0
12	0	1	1	1	0	0.61	0	0.61	0.61	0
13	0	1	1	1	1	0.56	0	0.56	0.56	0
14	0	1	1	0	0	0.5	0	0.5	0.5	0
15	0	1	1	0	1	0.4	0	0.4	0.4	0

Adapted from Mello (2014, 89); M: Military Power, V: Parliamentary Veto, C: Constitutional Restrictions, S: Public Support, E: Right Executive; MP: Outcome Military Participation.

However, the x-y plots of rows 6 and 8, respectively depicted in figures 2 and 3, do not indicate that the distribution of the cases of row 6 is more consistently in line with the distribution of a sufficient condition. Although only one inconsistent case is situated in row 6, its relatively high membership score of 0.63 makes it more relevant than the four inconsistent observations of row 8; none of which has a score above 0.5. Furthermore, the total sum of the inconsistent parts of these four inconsistent cases amounts to 0.44, only slightly exceeding the 0.43 inconsistency displayed by the disconfirming case of row 6. In line with Ragin's second requirement for the consistency measure, row 6 thus displays a "more glaring inconsistency". While the size of the inconsistencies of both rows are roughly equal, the inconsistencies of row 6 are caused by a relatively large, and thus more relevant, inconsistent case. Both rows have only one relatively large consistent case, but membership scores of the consistent cases are generally larger in row 6. This however does not justify a difference of 0.07 in their consistency scores, since the inconsistency of row 6 can be attributed to more relevant cases.

FIGURE 2: X-Y PLOT ROW 6 MELLO

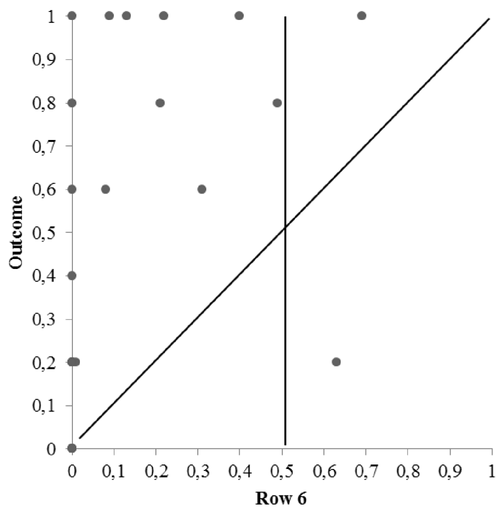
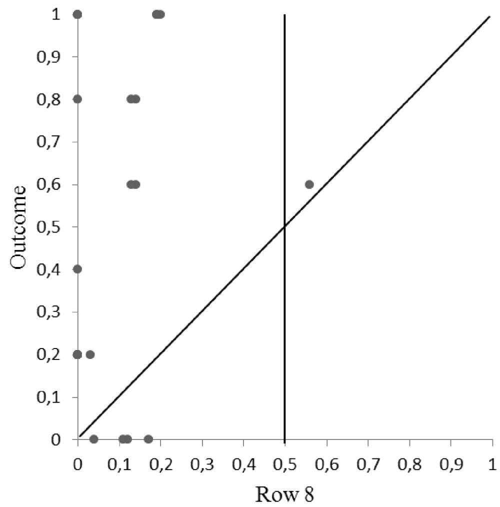
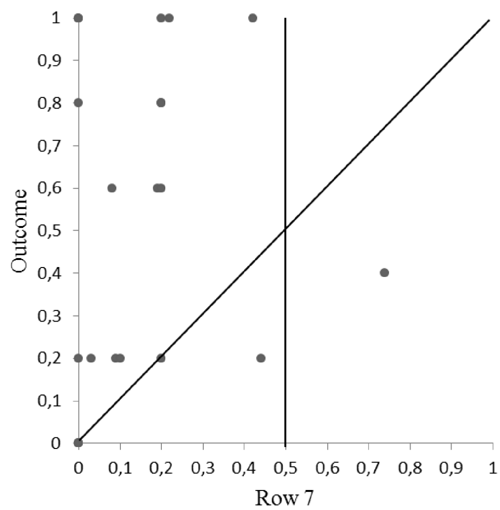


FIGURE 3: X-Y PLOT ROW 8 MELLO



When looking at the distribution of the cases in the outcome and rows 6 and 8, it seems appropriate to code both as sufficient for the outcome. However, this would require deciding on a consistency threshold below the consistency of row 8. Hereby, row 7 would also be coded sufficient. Although its consistency score exceeds the score of row 8 by 0.03, the sufficiency of row 7 is more flagrantly contradicted by the distribution of the cases –as demonstrated by its x-y plot presented in figure 4. There are two inconsistent cases in row 7, both of which are more relevant than the cases of row 8. With a score of 0.74, the first strongly exceeds 0.5. The other inconsistent case has a smaller membership score, but with a score of 0.44 still exceeds the largest inconsistent case of row 8. Furthermore, the sum of the inconsistent parts of these disconfirming observations amounts to 0.58, 0.14 more than the sum of the inconsistent parts of row 8. On top of that, none of the scores of the consistent cases exceeds 0.5, indicating that none of them can be considered relatively large. Nevertheless, because the consistency measure ascribes more substantial penalties for small inconsistent cases, it indicates that the empirical evidence more consistently confirms the sufficiency of row 7.

FIGURE 4: X-Y PLOT ROW 7 MELLO



Whereas the empirical evidence for the sufficiency of row 8 is thus roughly as robust as the evidence for row 6 and clearly stronger than the evidence for row 7, it has a considerably lower consistency score. However, the standard consistency measure does not allow to code row 8 as sufficient without coding the more inconsistent row 7 as insufficient. In order to alleviate this problem, consistency was calculated with the first adjusted formula, which does ascribe higher penalties for larger inconsistent cases. Whereas the score of row 8 remained constant at 0.8, the consistency of row 6 and 7 dropped to respectively 0.81 and 0.64. Rows 6 and 8 are hereby clearly set apart from the more inconsistent row 7. As argued above however, this alternative formula does not take into account the size of the inconsistencies.

In order to accommodate this shortcoming, the new consistency measure was used to assess the consistency of the truth table rows. The resulting score of row 8 exceeds the score of row 7, while leaving the order of the other rows largely unchanged. Hereby, the consistency threshold can be established below row 8, which is the 7th row of the new truth table. In consequence, rows 6 and 8 are considered sufficient for the outcome, whereas row 7 is coded as insufficient. The alternative consistency scores and the resulting outcomes are presented on the right-hand side of table 2.

The alternative coding of the outcome has a significant impact on the results of the analysis. This is demonstrated in table 3, which sets the resulting solutions against Mello's original solutions. Each row of this table corresponds to a specific sufficient combination for the outcome. The conditions are expressed in capital letters, a tilde refers to the absence of a condition and multiplication to the combination of conditions. The first path of the original formula thus refers to the presence of military power (M) in combination (*) with the absence of constitutional restrictions (~C). When comparing the solutions of Mello's analysis and the new analysis, two main differences appear. First, the first path of the parsimonious solution of the reanalysis does not include parliamentary veto power. Second, a new sufficient pathway appears in both the intermediate and complex solution.

TABLE 3: SOLUTIONS MELLO (2014)

	Original Solution	New Solution
Parsimonious Solution	M***~C**	M
	~V*S	~V*S
	~V*E	~V*E
Intermediate Solution	M*~C*E	M*~C*E
	~V*S*~E	~V*S*~E
	~V*~C*E	~V*~C*E
	/	**M*S*~E**
Complex Solution	M*V*~C*~S*E	M*V*~C*~S*E
	~M*~V*S*~E	~M*~V*S*~E
	~V*~C*S*E	~V*~C*S*E
	~M*~V*~C*E	~M*~V*~C*E
	/	**M*V*C*S*~E**

Adapted from Mello (2014, 90); ~ indicates absence of condition, * conjuction of conditions; differences between solutions are set in bold.

In order to compare the available evidence for both solutions, the x-y plots of the intermediate solutions of both analyses are depicted in figure 5 and 6. These reveal that the new intermediate solution covers four additional inconsistent cases. These correspond to the inconsistent cases of row 8, none of which has a membership score that exceeds 0.5. The new intermediate formula however also covers an additional large confirming observation. This seems to outweigh the downside of including the four small inconsistent cases, since the new formula hereby covers twelve of the thirteen democracies that participated in the operation in Kosovo.

FIGURE 5: X-Y PLOT INTERMEDIATE SOLUTION MELLO

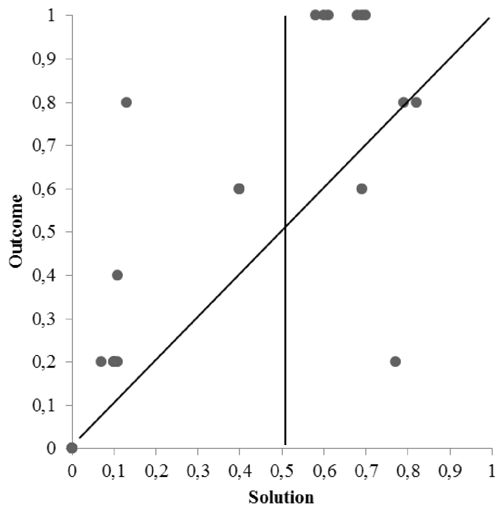
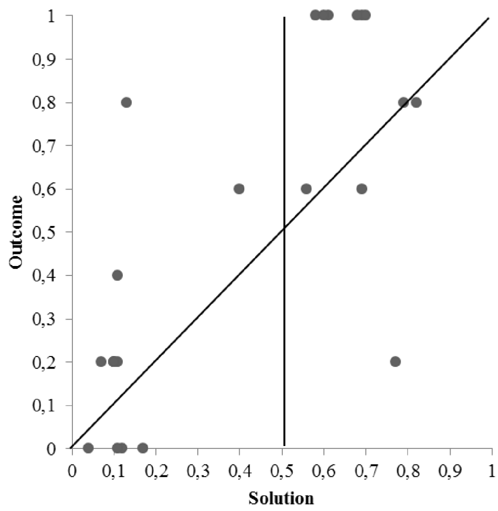


FIGURE 6: X-Y PLOT ALTERNATIVE SOLUTION MELLO



The alternative coding of the truth table rows has considerable implications for the conclusions of the analysis. First, Mello (2014, 90) concluded from the parsimonious and intermediate solution that parliamentary veto rights or constitutional restrictions, which both act as domestic institutional constrains on military deployment, needed to be absent for military participation. However, the first path of the new parsimonious solution indicates that military power is sufficient, independent of such institutional constrains. Similarly, the additional pathway included in the intermediate solution indicates that in the presence of military power, non-right executives participate if there is public support for military engagement, irrespective of institutional constraints. Second, the latter combination also confirms Mello's assertion that the conditions under which right and left leaning executives decide on military participations diverge, but allows for more fine-cut conclusions. More specifically, the alternative formula clearly reveals that public support is crucial for the participation of left-leaning executives, while the absence of constitutional restrictions is more important for right-leaning executives.

4.2 Schneider and Makszin: Welfare capitalism and participatory inequality

The second example is drawn from an article by Schneider and Makszin (2014), in which the authors assess whether a country's level of political inequality is shaped by features of its welfare system. Carsten Schneider can be considered one of the leading QCA experts. Not only does he teach the QCA course in the ECPR summer school, he also published methodological work on QCA and is co-author of one of the standard QCA textbooks (Wagemann and Schneider 2010; Schneider and Wagemann 2012; Schneider and Rohlfing 2013). As could be expected, the QCA-application in this example is flawless, so the misinterpretation of the evidence can be fully attributed to the flaw in the consistency measure.

The goal of the fsQCA was to unravel which (combinations of) welfare state characteristics cause participatory inequality. Four welfare capitalist features were included in the analysis: employment protection (EPL), labour market expenditure (LMX), wage coordination (WC) and union density (UD). 77 cases were included in the analysis, resulting in the truth table presented on the left hand side of table 4. The consistency threshold was set at 0.83. Consequentially, row 1-7 were coded sufficient, row 8-15 insufficient.

TABLE 4: TRUTH TABLE SCHNEIDER AND MAKSZIN (2014)

	Conditions				Original Analysis		New Analysis		
row	LMX	WC	UD	EPL	Standard Consistency	LPI	Previous Consistency	Adjusted Consistency	LPI
1	1	0	0	1	0.94	1	0.94	0.94	1
2	1	0	1	1	0.93	1	0.86	0.91	1
3	1	1	0	1	0.89	1	0.88	0.89	1
4	1	1	1	0	0.87	1	0.81	0.85	1
5	0	0	0	1	0.85	1	0.75	0.84	1
6	1	1	1	1	0.84	1	0.61	0.79	0
7	0	1	0	1	0.84	1	0.81	0.82	1
8	0	0	1	1	0.81	0	0.6	0.76	0
9	0	1	1	1	0.81	0	0.68	0.79	0
10	0	1	1	0	0.8	0	0.75	0.78	0
11	1	0	1	0	0.8	0	0.76	0.78	0
12	0	0	1	0	0.79	0	0.69	0.75	0
13	1	0	0	0	0.77	0	0.62	0.73	0
14	0	1	0	0	0.69	0	0.61	0.67	0
15	0	0	0	0	0.58	0	0.48	0.56	0

Adapted from Schneider and Makszin (2014, 450) LMX: High Labour Market Expenditure, WC: High Wage Coordination, UD: High Union Density, EPL: High Employment Protection; LPI: Outcome Low Participatory Inequality.

However, the x-y plots of row 6 and row 10, respectively depicted in figure 7 and 8, do not indicate that the distribution of the cases of the former is more consistently in line with the distribution of a sufficient condition. There are seven inconsistent cases in row 6, four of which have a membership score above 0.5. In contrast, six cases are inconsistent with the statement that row 10 is sufficient for the outcome, of which only one has a fuzzy score that exceeds 0.5. On top of that, the sum of the inconsistencies of these six cases amounts to 1.05. This is significantly less than the sum of the inconsistencies of row 6, which amounts to 1.25. Row 6 thus displays larger inconsistencies, which can be attributed to cases with larger membership scores. Row 10 furthermore includes three consistent cases that exceed the 0.5 threshold, whereas row 6 only includes two relatively large consistent cases. Nevertheless, the consistency measure indicates that row 6 is more consistent than row 10.

FIGURE 7: X-Y PLOT ROW 6 SCHNEIDER AND MAKSZIN

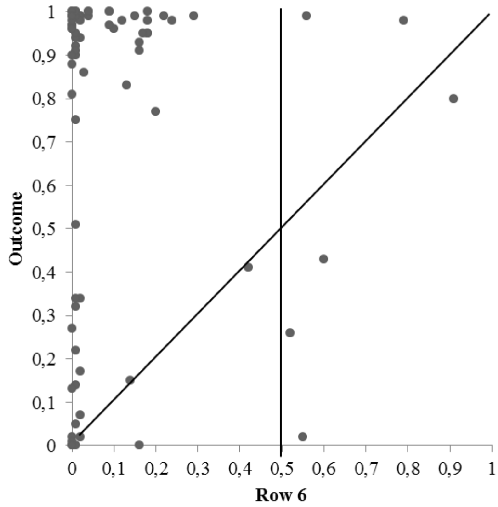
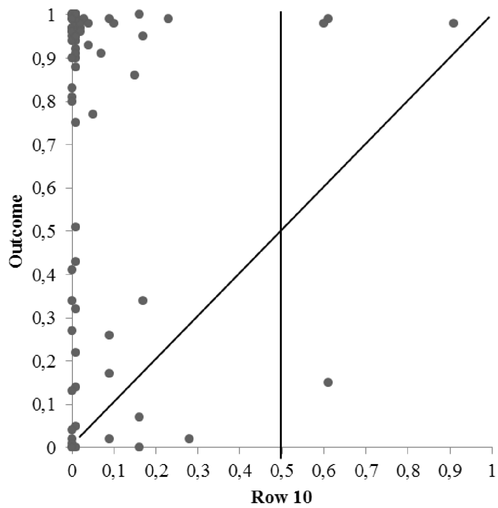
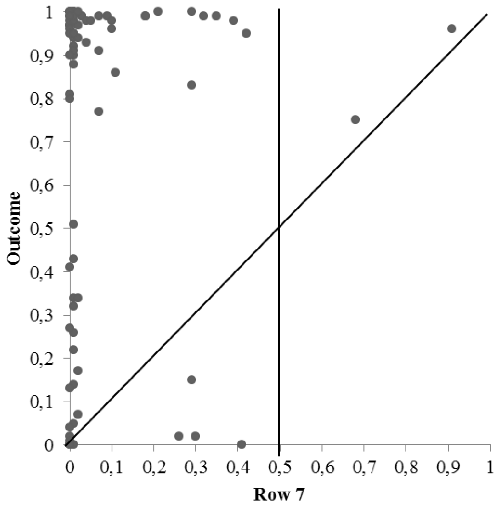


FIGURE 8: X-Y PLOT ROW 10 SCHNEIDER AND MAKSZIN



When looking at the distribution of the cases in the outcome and row 6 and 10, it seems more appropriate to code both rows as insufficient. However, this would require establishing a consistency threshold above the consistency of row 6. Hereby, row 7 would also be coded as insufficient. Although its consistency falls slightly below the score of row 6, the cases more consistently support the sufficiency of row 7. The x-y plot, depicted in figure 9, demonstrates that there are only 5 inconsistent cases in row 7, of which none has a membership score above 0.5. On top of that, the sum of the inconsistent parts of these cases amounts to 1.08, considerably less than the sum of the inconsistencies of row 6. Since both row 6 and 7 have two consistent cases with a score above 0.5, the higher consistency of the former can clearly be attributed to the fact that small inconsistent cases have more impact than large inconsistent cases.

FIGURE 9: X-Y PLOT ROW 7 SCHNEIDER AND MAKSZIN



Whereas the empirical evidence for the sufficiency of row 6 is thus as weak as the evidence for row 10, its consistency is significantly higher. Similarly, although the distribution of the membership scores in row 7 more consistently confirms a subset relation, its consistency falls slightly below the consistency of row 6. Consequentially, the standard consistency measure does not allow to code row 6 and 10 as insufficient, while coding 7 as sufficient. To alleviate this problem, consistency was calculated with the first adjusted formula. This results in a significantly lower score for row 6, whose consistency dropped from 0.84 to 0.61. In contrast, the consistency of row 7 only dropped from 0.84 to 0.81, the score of row 10 from 0.8 to 0.75. As argued above however, this alternative formula does not take into account the size of the inconsistencies.

In order to accommodate this shortcoming, the new consistency measure was used to assess the consistency of the truth table rows. The resulting score of row 7, which is the 6th row of the new truth table, still exceeds the score of row 6 and 10. Since the consistency scores drop considerably after row 7, the cut-off point can be established just below its consistency. Hereby, row 7 is coded sufficient for the outcome, row 6 and 10 insufficient. The alternative consistency scores and the resulting outcomes are presented on the right-hand side of table 4.

The new coding of the outcome value has a significant impact on the results of the analysis. In line with Schneider and Makszin's original analysis, the logical remainder is included in the analysis. Two of the three sufficient combinations display an additional condition in the resulting solution: the first includes the absence of wage coordination, the second the absence of employment protection. The resulting formula is set against the original solution in table 5.

TABLE 5: SOLUTION SCHNEIDER AND MAKSZIN (2014)

	Original Solution	New Solution
Solution	EPL*LMX	EPL*LMX***~WC**
	LMX*WC	LMX*WC***~EPL**
	EPL*~UD	EPL*~UD

Adapted from Schneider and Makszin (2014, 452); ~ indicates absence of condition, * conjuction of conditions; differences between solutions are set in bold.

The x-y plots of the original and new solutions, respectively presented in figures 10 and 11, are used to compare the evidence for the solutions. These reveal that the new solution covers seventeen inconsistent cases, of which seven are larger than 0.5. In contrast, the old solution covered nineteen inconsistent cases, of which eleven were larger than 0.5. This loss of four large inconsistent cases only comes at the cost of losing two large consistent cases. Since the new solution still covers twenty-two large consistent cases, the ratio inconsistent to consistent observations more strongly confirms the sufficiency of the new solution.

FIGURE 10: X-Y PLOT SOLUTION SCHNEIDER AND MAKSZIN

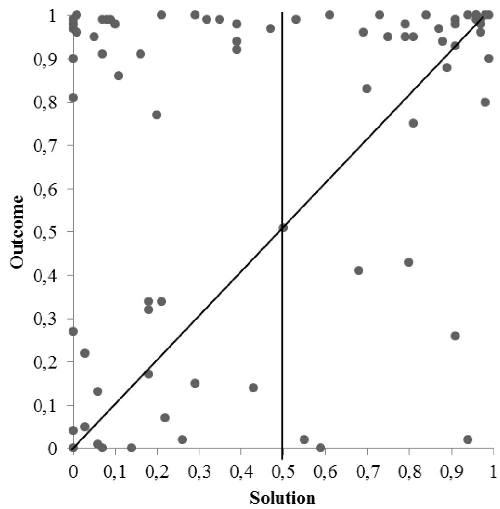
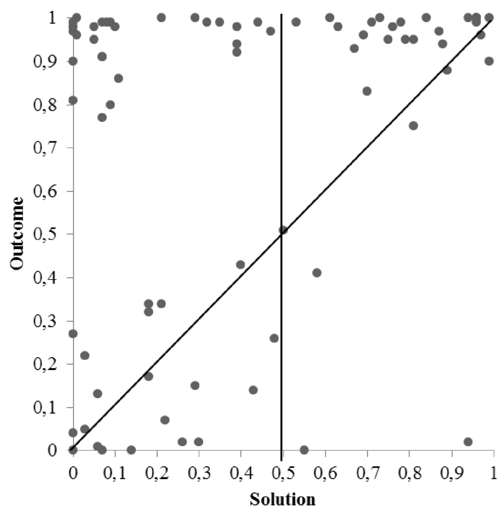


FIGURE 11: X-Y PLOT ALTERNATIVE SOLUTION SCHNEIDER AND MAKSZIN



The alternative coding of the truth table rows has considerable implications for the conclusions of the analysis. Schneider and Makszin (2014, 452) concluded from their results that employment protection and wage coordination are both sufficient in combination with high level labour market expenditure and thus act as functional equivalents. However, the new results indicate that only one of these welfare state treats can be present to allow for low participatory inequality. Evidently, this induces different conclusions on the impact of welfare state characteristics on participatory equality.

5 CONCLUSION

The consistency measure constitutes QCA's single most important parameter for assessing sufficient and necessary conditions. This article however revealed that the currently used formula does not accurately express the degree to which empirical evidence supports statements of sufficiency and necessity. Because small inconsistent cases exert more impact than large inconsistent cases, the currently used consistency measure can lead to the misinterpretation of empirical evidence for sufficiency and/or necessity. As demonstrated by the re-analysis of two recent QCA-applications, this has significant implications for the conclusions of their research project. To alleviate the problems of the standard consistency measure, this article introduced a new formula for calculating consistency, which more accurately assesses the evidence for sufficiency and/or necessity.

REFERENCES

Marx, Axel, Benoît Rihoux and Charles Ragin. 2014. "The origins, development, and application of Qualitative Comparative Analysis: the first 25 years." *European Political Science Review* 6 (1): 115–142.

Mello, Patrick A. 2012. "Parliamentary Peace or Partisan Politics? Democracies' Participation in the Iraq War." *Journal of International Relations and Development* 15 (3): 420–453.

Mello, Patrick A. 2013. " From Prospect to Practice: A Critical Review of Applications in Fuzzy-Set Qualitative Comparative Analysis." *Paper presented at the 8th Pan-European Conference on International Relations, Warsaw, 18-21 September*.

Mello, Patrick A. 2014. *Democratic Participation in Armed Conflict Military Involvement in Kosovo, Afghanistan, and Iraq*. Basingstoke: Palgrave Macmillan.

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*. Chicago: University of Chicago Press.

Ragin, Charles C. and Sean Davey. 2012. *fs/QCA [Computer Programme]* Irvine, CA: University of California.

Rihoux, Benoît. 2014. Compasss Newsletter (33). Available at http://www.compasss.org/newsletter.htm.

Rihoux, Benoît and Charles C. Ragin. 2009. *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*. London: Sage Publications.

Schneider, Carsten Q. and Bernard Grofman. 2006. "It Might Look like a Regression Equation... But it's Not! An Intuitive Approach to the Representation of QCA and FS/QCA Results." *COMPASSS WP Series* 2006 (39): 1–61.

Schneider, Carsten Q. and Kristin Makszin. 2014. "Forms of welfare capitalism and education-based participatory inequality." *Socio-Economic Review* 12 (2): 437–462.

Schneider, Carsten Q. and Ingo Rohlfing. 2013. "Combining QCA and Process Tracing in Set-Theoretic Multi-Method Research." *Sociological Methods & Research*.

Schneider, Carsten Q. and Claudius Wagemann. 2012. *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis (QCA)*. Cambridge: Cambridge University Press.

Thiem, Alrik and Adrian Dușa. 2013. *Qualitative Comparative Analysis with R: A User's Guide*. New York: Springer.

Wagemann, Claudius and Carsten Q. Schneider. 2010. "Qualitative Comparative Analysis (QCA) and Fuzzy-Sets: Agenda for a Research Approach and a Data Analysis Technique." *Comparative Sociology* 9 (3): 376–396.