Van der Heijden, J. (2014) Voluntary programmes for building retrofits: opportunities, performance and challenges, *Building Research & Information*. Doi: 10.1080/09613218.2014.959319

SUPPLEMENTAL FILE

Applying fuzzy set qualitative comparative analysis (fsQCA) in this study

QCA was introduced by Charles Ragin, a social scientist, as a middle path between quantitative and qualitative social research (Ragin, 1987). QCA is grounded in set theory, a branch of mathematical logic that allows studying in detail how causal conditions contribute to a particular outcome. QCA has since the mid-1990s quickly evolved as an accepted research practice for the type of study presented in this article, and has been applied in hundreds of studies in the policy sciences in particular and the social sciences more generally (Rihoux, 2013; Rihoux & Ragin, 2009). See for instance a recent special issue of *Policy and Society* on 'Innovative methods for policy analysis: QCA and fuzzy sets' (Policy and Society, 2013).

Particularly the introduction of fsQCA appears to have spurred the use of the method by a range of scholars from various backgrounds studying issues such as governance networks within a country (Verweij, Klijn, Edelenbos, & Van Buuren, 2013), job security regulations in Western democracies (Emmenegger, 2011), or organisational configurations (Fiss, 2011). In a recent article, Barbara Vis (2012) has compared fsQCA with Regression Analysis by applying the two data analysis techniques to a single dataset. She finds that fsQCA is better capable to understand and make visible complex relations than regression analysis (for a comparable argument, see Warren, Wistow, & Bamba, 2013).

The fundamentals and background of the method are well-explained and documented in a series of strong textbooks (Goertz & Mahony, 2012; Ragin, 2008; Rihoux & Ragin, 2009; Schneider & Wagemann, 2012). These handbooks are good further references for those unfamiliar with the foundations of the method (which I will not dwell on here).

These handbooks provide guidelines for good fsQCA practice (Ragin, 2008, see in particular the 'practical appendices'; Rihoux & Ragin, 2009, particularly Chapter 5; Schneider & Wagemann, 2012, Chapter 11), which were closely followed in undertaking the analyses discussed in the article. One of the key points for good fsQCA practice is for the researcher to provide as much transparency into the analysis as possible. This supplementary appendix seeks to provide this.

This appendix provides an account of the various steps taken and decisions made in the fsQCA analysis as presented in the article. Where necessary, additional (raw) data is provided so that the interested reader can repeat this analysis. In this account, the 'flowchart' of Jerry Mendel and Mohammad Korjani (2013) is followed. These authors, supported by Charles Ragin, have mathematically summarized fsQCA as a collection of 13 steps. I do however take the liberty to use the jargon from the handbooks (as compared to the mathematical jargon introduced by Mendel and Korjani) to specify the steps and condense them into 10 steps.

In addition to Mendel and Korjani's steps of *how* the fsQCA analysis is carried out it is, of course, of importance to motivate *why* fsQCA was chosen in the first place. Whilst researchers often support their choice for fsQCA with a practical motivation (*i.e.* they have a medium number of cases that likely allows for systematic cross-case analysis, but not for sophisticated statistical analysis), ideally fsQCA is chosen for a theoretical motivation (Schneider & Wagemann, 2012). I have added this step of motivating the choice for fsQCA, and will start with it below.

Step 1: Why an fsQCA analysis?

Earlier empirical studies have found that, for instance, voluntary programmes with similar designs such as pay-per-plastic-bag fees (Ackerman, 1997), organic food labelling (Thøgersen, 2010), building assessment classification and certification (Fowler & Rauch, 2006), and revolving loan funds (Boyd, 2013) show different outcomes depending on how their design conditions interact with contextual conditions (*e.g.* existing legislation, economic circumstances; Borck & Coglianese, 2009). Even more, some studies indicate that a single design (*e.g.* building assessment classification and certification) implemented in a number of similar contexts (*e.g.* US, AU, UK) may nevertheless result in different outcomes due to the role of governmental actors in these arrangements (Fowler & Rauch, 2006).

This all indicates that the outcomes of voluntary programmes are likely caused by different interacting conditions (*i.e.* conjunctural causation), that different (configurations of interacting) conditions may cause a similar outcome (*i.e.* equifinality), and that the presence of a (configuration of interacting) condition(s) in the causal role of the outcome is of limited help in explaining the inverse situation. (In other words, the causal role of the absence of the condition in the non-occurrence of the outcome, *i.e.* an asymmetry).

QCA is chosen as a data analysis methodology because it allows for 'unraveling causally complex patterns in terms of equifinality, conjunctural causation, and asymmetry' (Schneider & Wagemann, 2012, 8). QCA differs from other data analysis methods in its focus. 'The key issue [for QCA] is not which variable is the strongest (*i.e.* has the biggest net effect) but how different conditions combine and whether there is only one combination or several different combinations of conditions (causal recipes) of generating the same outcome' (Ragin, 2008, 114). QCA helps to trace patterns of association between these conditions in a highly systemised manner and allows for systematic comparisons between empirical observations (*i.e.* cross-case), whilst allowing for in-depth within-case understanding of the individual observations (Rihoux & Ragin, 2009).

I have chosen fuzzy set QCA (fsQCA) as it allows for giving a rather precise insight in the qualitative difference in my empirical data – *i.e.* the degree of presence or absence of a condition or the outcome in the cases under analysis. This particular issue is explained in more depth under step under step 4.

Step 2: Selection of outcome of interest and cases to study

In the article, the outcomes of interest are explained. The operationalization of these outcomes is further explained below. The selection of cases (real world examples of innovative governance voluntary programmes)

is also explained in the article. In short, I have selected 20 cases of voluntary programmes from a larger study (Van der Heijden, forthcoming, 2014).

Step 3: Select k causal conditions

The article discusses a set of seven design conditions that the current literature considers most relevant in influencing the performance of voluntary programmes. The operationalization of these conditions is explained below.

Step 4: Calibration of set-membership scores for outcomes and conditions

The strength of fsQCA as compared to other forms of QCA is that it allows for giving a rather precise insight in the qualitative difference in the units of observation. In other words, it allows distinguishing among different qualitative categories of these observations and compare sets of observations of a particular category with sets of observations of other categories.

To illustrate, imagine that you have to classify the greenness (colour) of 10 paintings for a particular analysis. You could, of course, measure the percentage of green in the paintings and rank them accordingly, but how to deal with different shades of green? Or, is a fully green 2"x2" painting greener than a 3'2"x3'2" that has 'only' 50% of its canvas painted green? And what about a painting that is predominantly yellow and blue, the prime colours that together make green?

Before starting this seemingly easy task, you will have to create at least two categories for ranking the paintings: one for the paintings that meet a certain understanding of green, and the others that do not. Assume that you decide that in order to be considered 'green' half of a painting needs to be painted a shade of green. This categorisation results in seven paintings meeting the condition green, and three paintings that do not. Upon second inspection you find that two of those seven in the "in" category are very green (say, 80% or more); and one of the three paintings in the "out" category has no green at all, leaving two paintings somewhat green. In analysing the paintings it may be of interest to use these qualitative differences in the original "in" and "out" categories. You therefore decide to group those two paintings from the original "in" category as having "full-membership" in the condition green and the five remaining paintings as being "more in than out" of the condition green. The two somewhat green paintings in the original "out" category could be further grouped as "more out than in" the condition green, leaving the last painting in the category "full non-membership" in the condition green.

This is precisely what calibration of data in fsQCA implies. It asks the researcher to carefully distinguish the various qualitative categories of their observations according to their qualitative differences and carefully assign their data to these categories. Good fsQCA practice requires the researcher to be clear about this calibration. Particularly to explain the two extremes of the observed data (*i.e.* maximum and minimum fit in a category), and the crossover point of the data (*i.e.* in what stage is the data considered to have maximum ambiguity; that is, when is it as much in as out?) (Ragin, 2008; Rihoux & Ragin, 2009;

Schneider & Wagemann, 2012). I have calibrated my data using a four category qualitative scale as represented in table A1.

Table A1: Verbal description of membership scores of the data in qualitative categories

The observation is...	Qualitative symbol	Fuzzy set value
Full membership (<i>i.e.</i> in the highest stage observed)	++	1.00
More in than out	+	0.67
More out than in	-	0.33
Full non-membership (<i>i.e.</i> in the lowest stage observed)	--	0.00

For the various outcomes and conditions the extremes and crossover points in the data are set as follows:

1. Observed outcome:
 - **Participants (O1).** I have operationalised this outcome by considering the participants a voluntary programme has attracted as compared with stated ambitions (in documentation or expressed in interviews). Full membership represents that stated ambitions are met; full non-membership represents not having attracted any or only a few participants; and, the crossover point is set at not meeting half the stated ambitions in terms of attracting participants.
 - **Building retrofits (O2).** I have operationalised this outcome by considering the number of building retrofits achieved under a voluntary programme as compared with stated ambitions (in documentation or expressed in interviews). Full membership represents that stated ambitions are met; full non-membership represents not having achieved any building retrofits at all; and, the crossover point is set at not meeting half the stated ambitions in terms of building retrofits.
2. Conditions:
 - **Financial gain (Fg).** The qualitative categories for the direct financial gain (including cost savings) participants may get from joining an voluntary programme and retrofitting buildings are constructed by combining data on 'promised' gains (*i.e.* how prospective gains are marketed by the administrators of these voluntary programmes) and 'evidenced' gains (*i.e.* how realised gains are marketed by administrators and participants of these voluntary programmes). Full membership represents a marketed high certainty of achieving substantial financial gains when participating based on evidence. More in that out membership represents a marketed promised certainty of gains supplemented with evidence. More out than in represents a marketed promise of gains when participating. Full non-membership represents a full absence of a marketing of gains. The crossover point is the marketing of promised high certainty of gains but without evidence to support this promise.
 - **Non-monetary gain (Nm).** I followed the above line of reasoning. Thus, full membership represents a marketed high certainty of achieving substantial non-monetary gain when participating based on evidence; full non-membership represents a full absence of a marketing of non-monetary gains; and,

the crossover point is the marketing of promised high certainty of non-monetary gains but without evidence to support this promise.

- **Showcasing leadership (Le).** To construct a fuzzy set for this condition I considered how administrators of voluntary programmes reward and market leadership. Full membership represents a focus on national or global leadership combined with marketing of leading practice or awarding of leading practice through, for instance, yearly awarding ceremonies. More in than out represents a focus on regional or local leadership combined with marketing or awarding of such leadership. More out than in represents a focus on leadership in the marketing of a voluntary programme, but an absence of marketing or rewarding actual leadership by participants. Fully non-membership represents a full absence of a focus on leadership in the marketing of a voluntary programme. The crossover point of this condition is the marketing of best-practices as opposed to local, national or international leadership.
- **Participation criteria (Pc).** I followed the above line of reasoning. Thus, full membership represents that participants are required to perform significantly beyond the requirements of public law and regulation (*e.g.* to achieve double the statutory requirement, or to show high level performance in an area that is not yet addressed through statutory regulation). More in than out represents that participants are required to perform well beyond the requirements of public law and regulation (*e.g.* to achieve more than the statutory requirement, or to show unspecified performance in an area that is not yet addressed through statutory regulation). More out than in represents that participants are required to perform just beyond the requirements of public law and regulation.¹ Full non-membership represents a full absence of criteria. The crossover point of this condition is set at criteria that only require performance that is marginally better than what is required by statutory regulation.
- **Enforcement criteria (En).** The qualitative categories reflect the strictness of enforcement in terms of who enforces, how enforcement is carried out, and what evidence results from enforcement. Full-membership represents strict enforcement; for instance third-party enforcers, a documented enforcement process, and the awarding of a certificate at the end of the process. More in than out represents medium enforcement; for instance, administrator enforcement and documented proof of compliance at the end of the process. More out than in represents a weak enforcement process; for instance, participant self-enforcement. Full non-membership represents the absence of enforcement. The crossover point of this condition is set at administrator enforcement without documented proof of compliance at the end of the enforcement process.
- **Government involvement (Gi).** To construct a fuzzy set for this condition I have considered how governments are involved in the voluntary programmes. Full membership represents sole governmental involvement in initiating and administering a voluntary programme. More in than out represents sole governmental involvement in initiating a voluntary programme. More out than in represents equal involvement in the initiating of voluntary programmes of governmental and non-

¹ A voluntary programme may seek to ease or stimulate compliance with statutory regulation.

governmental actors. Full non-membership represents the absence of government involvement. The crossover point of this condition is set at dominance of governmental involvement in this role.

- **Government participation (GP).** Full membership represent high activity; for instance, government as dominant participants or customers of an voluntary programme as a result of mandatory participation or procurement. More in than out represents medium activity; for instance, mandatory participation or procurement criteria (but no government dominance in an voluntary programme's participants or customers). More out than in represents low activity; for instance, preferred participation or procurement. Full non-membership represents no government role. The crossover point is set at specified requirements for governments to participate in, or require their suppliers to participate in specific voluntary programmes.

Step 5: Create a raw data matrix

Now that the various qualitative differences of the outcomes and conditions have been distinguished the data can be transformed into a raw data matrix. Table A2 provides the raw data used in the article.

Table A2: Raw data matrix

Case*	Condition**							Outcome**	
	<i>Fi</i>	<i>Nm</i>	<i>Le</i>	<i>Pc</i>	<i>En</i>	<i>Gi</i>	<i>Gp</i>	<i>O1</i>	<i>O2</i>
1a	1.00	0.67	1.00	0.67	1.00	0.33	0.67	1.00	0.33
1b	1.00	0.67	1.00	1.00	1.00	0.33	0.67	1.00	0.00
4	0.67	1.00	0.00	0.33	0.00	0.67	0.00	0.67	0.33
6	1.00	0.33	0.33	0.67	0.33	1.00	0.67	0.67	0.00
9	1.00	0.33	1.00	1.00	0.67	1.00	0.33	0.67	1.00
14	1.00	0.00	0.00	0.67	0.67	1.00	0.33	0.33	0.33
21	0.67	0.33	0.00	0.00	0.00	0.67	0.33	0.33	0.00
22	1.00	0.67	0.67	0.67	0.67	0.00	0.00	0.67	0.33
23	1.00	0.33	0.33	0.33	0.33	0.33	0.33	1.00	0.67
24	0.67	0.33	0.33	0.67	0.67	1.00	0.33	0.33	0.33
25	0.33	0.67	0.00	0.67	0.67	1.00	0.67	1.00	0.67
26	0.33	0.67	0.67	0.67	0.33	0.33	0.67	1.00	0.33
27a	1.00	0.67	1.00	0.67	0.67	0.33	0.67	0.67	0.00
27b	1.00	0.67	1.00	1.00	0.67	0.33	0.67	0.67	0.00
42	0.67	0.33	0.00	0.67	0.67	1.00	0.33	0.33	0.00
46	0.67	0.33	0.00	0.33	0.67	0.67	0.00	1.00	0.67
49	1.00	0.67	0.67	0.67	0.00	0.67	0.00	0.33	0.33
50	0.67	0.33	0.00	0.33	0.67	0.33	0.00	0.33	0.33
54	1.00	0.67	0.67	0.67	0.33	0.67	0.00	1.00	0.67
55a	1.00	0.67	1.00	0.67	0.67	0.00	0.67	1.00	0.33
55b	1.00	0.67	1.00	1.00	0.67	0.00	0.67	1.00	0.00
63	0.67	0.67	0.00	0.67	0.33	0.00	0.00	0.33	0.00
64	1.00	0.33	0.33	0.67	0.33	1.00	0.67	0.33	0.00

Notes:

* Cases are given numbers to maintain anonymity as requested by some interviewees. Please note the 'a' and 'b' cases refer to specific arrangements that allow their participants to meet either high or moderate participation criteria.

** Conditions and outcomes as described in the article.

1.00 = maximum score (e.g. the arrangement has attracted at least the expected number of participants); 0.67 = score closer to “1.00” than to “0.00” (e.g. the arrangement has attracted a substantial number of participants, but not the expected number); 0.33 = score closer to “0.00” than to “1.00” (e.g. the arrangement has attracted a marginal number of participants, but this number far from meeting the expected number); 0.00 = minimum score (e.g. the arrangement has not attracted any or only a few participants)

Step 6: Analysis of necessary conditions

Following established fsQCA practice the data is first analysed for necessary conditions before exposing it to more complex analysis to identify (configurations of) sufficient conditions (Rihoux & Ragin, 2009, Chapter 5, box 8.1; Schneider & Wagemann, 2012, Chapter 11).

For a condition to be necessary to cause the outcome the fuzzy-set membership scores of the outcome need to be a perfect subset of the membership scores of the condition. To give an illustration, in order for it to be true that one can practice as an architect in the Netherlands (outcome) it is necessary that one is registered with the Netherlands Architects Registrar (condition). In other words, the set of architects in the Netherlands (outcome) is a subset of individuals registered with the Netherlands Architects Registrar (condition). The set of individuals registered with the Netherlands Architects Registrar is however (much) larger than the set of architects because it also includes landscape architects, city planners and interior designers.

To gain an insight as to whether any of the distinct conditions (see step 3) is necessary for causing the outcome (see step 3) I have plotted a series of fuzzy set XY plots using the computer program FS/QCA version 2.5 (cf., Schneider & Wagemann, 2012, Chapter 5 and Chapter 9). Table 4 in the article presents the results.

In studying necessity two issues are of importance: consistency and coverage. *Consistency* indicates how strongly the condition relates to the outcome. In other words, if a hypothesised relation between a condition and an outcome is not consistent (where the advisory cut-off point of consistency is a score of 0.90), the hypothesised relation cannot be supported by the data as being necessary (Rihoux & Ragin, 2009, 45). Table 4 in the article indicates that only the condition ‘financial gain’ passes the consistency test. However, the low coverage score of 0.35 indicates that this is likely a trivial necessary condition in achieving this outcome (Schneider & Wagemann, 2012, 232-237).

Coverage indicates how relevant the condition is for causing the outcome. Coverage is only assessed for conditions that meet the consistency test. Here it is important to distinguish between relevant and trivial necessary conditions. In other words, if a consistent relation only covers a small number of cases (*i.e.* if it has a low coverage score such as the condition ‘financial gain’) it can be considered to be trivial in causing the outcome (further, Schneider & Wagemann, 2012, Chapter 9). Another way to distinguish between relevant and necessary conditions is to assess whether or not the data is skewed towards conditions that have high scores for both the condition and the outcome. This suggests that such conditions may pass the test for both necessity and sufficiency, and is likely a trivial necessary condition in achieving this outcome (Schneider & Wagemann, 2012, 232-237).

Step 7: Analysis of sufficient conditions (1): create a truth table

Having studied the data for necessary conditions (but having found none) the next step is to study the data for sufficient conditions.

For a condition or for a configuration of conditions to be sufficient for causing the outcome the fuzzy-set membership scores of the condition or the configuration of conditions need to be a perfect subset of the membership scores of the outcome. To give an illustration, whilst being registered with the Netherlands Architects Registrar is a necessary condition for one to practice as an architect in the Netherlands this is however not a sufficient condition. After all, one also needs an office, the relevant design software, and so on, to be able to practice as an architect (*i.e.* other necessary conditions). In order to get the registration (outcome) it is however sufficient that one holds a Masters' Degree in Architecture from Delft University of Technology (condition), which the Registrar accepts as meeting the requirements for registration. This Degree is however not necessary for registration since the Registrar accepts degrees from other educational facilities as well (*i.e.* other sufficient conditions).

The analysis of (configurations of) sufficient conditions for the outcomes under scrutiny follows three sub-steps. The first sub-step is to create a truth table. Established QCA practice requires to present this truth table because it is the basis of the following analysis. Table A3 provides the truth table for the analysis of sufficient conditions for the outcome under scrutiny ('O2'). The truth table is created using FS/QCA software (version 2.5).

Table A3 - Truth table for the analysis of sufficient conditions for outcome O2

Row	Conditions							Number (case)	Outcome O2	Consistency Raw
	<i>FI</i>	<i>Nm</i>	<i>Le</i>	<i>Pc</i>	<i>En</i>	<i>Gi</i>	<i>GP</i>			
1	1	0	0	0	1	1	0	1	1	0.82
2	0	1	0	1	1	1	1	1	1	0.80
3	1	0	0	0	0	0	0	1	1	0.78
4	1	1	0	1	0	0	0	1	0	0.76
5	1	0	0	0	1	0	0	1	0	0.75
6	1	1	1	1	0	1	0	1	1	0.67
7	1	0	0	0	0	1	0	1	0	0.66
8	1	1	0	0	0	1	0	1	0	0.66
9	1	0	1	1	1	1	0	1	1	0.64
10	1	0	0	1	0	1	1	1	0	0.62
11	1	0	0	1	1	1	0	3	0	0.61
12	1	1	1	1	1	0	0	1	0	0.54
13	1	1	1	1	0	0	1	1	0	0.49
14	1	1	1	1	1	0	1	6	0	0.28
15-128	<i>Logical remainders</i>								?	

Note: The table and symbols are explained in the text below. Abbreviations as per table A2 and the article.

The truth table is a data matrix with 2^k rows that represents all possible configurations of conditions that are logically possible. Note, the truth table reports data using the cross-over points set – *i.e.* '1' indicates more in than out of the set (including full membership), and '0' indicates more out than in the set (including

full non-membership). Thus, with the seven conditions here the number of logically possible configurations is 128 (*i.e.* 2^7). The empirical observations are included in this table. As the truth table indicates, out of 128 logically possible configurations 14 were empirically observed (rows 1 to 14).

The different rows can be understood as ideal types (Schneider & Wagemann, 2012, Chapter 7). The number column ('No.') indicates how many cases fit best in this ideal type (*i.e.* when a case has a membership in the configuration of the fuzzy-sets for the conditions of at least 0.5). The row 'outcome' indicates whether for a configuration of conditions the outcome was observed or not (a '1' indicates it is, a '0' indicates it was not).

Because some observations of configurations of conditions may be observed in different cases, some rows in the truth table may refer to many cases (*e.g.* row 14) whilst other rows refer to only a few or just one case (*e.g.* rows 1 and 11). It is normal that the truth table also contains rows of possible combinations, but without empirical observations (*i.e.* rows 15 - 128).

In the second sub-step the truth table is logically minimized based on two conditions. First, the researcher sets a threshold for 'logical remainders'. Logical remainders are those configurations of conditions that 'lack enough empirical evidence to be subjected to a test of sufficiency' (Schneider & Wagemann, 2012, 152). It depends on the size of the research project (*i.e.* the number of cases included) what is to be considered as 'enough empirical evidence'. Most often a threshold of one observation (thus at least one case) is used, but for larger numbers of cases a higher threshold can be applied (Ragin, 2008; Schneider & Wagemann, 2012). Following this practice I have decided a threshold of at least one observation.

Second, the researcher has to set a 'consistency threshold for distinguishing [configurations of conditions] that are subsets of the outcome from those that are not' (Ragin, 2008, 143). In other words, how well do the configurations of conditions fit the outcome? This is what the 'raw consistency' score in the truth table indicates. As discussed under step 6, the higher the score the better the fit. Ragin (2008) advises a consistency score of at least 0.75, which I have followed. Please note, the consistency score of row 5 before rounding was < 0.75 .

This resulted in a minimization in which 4 cases met the observation threshold; and two cases met the consistency threshold for the outcome under scrutiny. In FS/QCA cases that met the consistency threshold were labelled '1' in the outcome column in the and those that did not were labelled '0' (cf., Ragin, 2008, 144).

Step 8: Analysis of sufficient conditions (2): dealing with logical remainders and choice of solution term

Having carried out this minimization of the truth table a standard analysis can be run in FS/QCA (the third sub-step). This standard analysis is best understood as the identification of 'the combinations of attributes [*i.e.* configurations of necessary conditions] associated with the outcome of interest using Boolean algebra and algorithms that allow logical reduction of numerous, complex causal [configurations of] conditions into a reduced set of configurations that lead to the outcome' (Fiss, 2011, 402). Normally a standard analysis results in a solution that consists of a number of 'paths' or 'solutions' (combinations of sufficient conditions) that lead to the outcome.

A simple example may explain what is going on in this analysis: imagine two painters painting a fully green painting. You are interested to see what colour green they are using. In inspecting their paint boxes you find that Painter A has yellow, blue and red paint; and Painter B has yellow and blue paint but not red. How then have they come to green?

First, they must have mixed their paints as blue, yellow, or red cannot suddenly become green (or at least, let us assume that for the sake of the example). Second, they have mixed at least two paints. But what mixes have they used?

Painter A could have mixed 'blue' and 'yellow' and 'red', whilst Painter B could have mixed 'blue' and 'yellow' but not 'red'. It logically follows that 'red' is not needed to make green. In other words, red can be eliminated as a condition needed for the outcome green, which leaves the combination 'blue and yellow' as the configuration of paints that causes green. This is in a nutshell what the logical reduction of causal configurations of conditions implies (more sophisticated explanations are found in the handbooks by Goertz & Mahony, 2012; Ragin, 2008; Rihoux & Ragin, 2009; Schneider & Wagemann, 2012).

The standard analysis produces three types of logically reduced configurations of conditions that are sufficient for the outcome under scrutiny: a complex solution, an intermediate solution and a parsimonious solution.

The complex solution is exclusively based on the empirical information at hand. The complex solution can however be further simplified by using counterfactuals for the logical remainders. Distinction is made in 'easy counterfactuals' and 'difficult counterfactuals' (this is well explained by Fiss, 2011). Easy counterfactuals are based on the theoretical assumptions (or other substantive knowledge by the researcher), for this study these are the assumptions identified in table A6 (above). Including easy counterfactuals in the standard analysis leads to the intermediate solution. Another illustration may be helpful here.

Assume again those two painters, but this time you cannot observe the yellow paint in Painter A's paint box (say, after mixing green Painter A gave her yellow to Painter B who forgot to bring it). If theoretical or otherwise substantive knowledge exist that in order to make green paint yellow paint is needed then the counterfactual 'yellow' may be added to the analysis. The intermediate solution would then (again) indicate that in order to get 'green' 'blue and yellow' are needed, and that 'red' can be eliminated as a condition.

The parsimonious solution (*i.e.* the most simplified solution) results from using difficult counterfactuals. Applying difficult counterfactuals is the inverse of applying easy counterfactuals. That is, assumptions are made about the outcome of a configuration if the counterfactual condition is redundant. This is a more complicated (and risky) undertaking since expectations are often made on conditions being present, and not absent. Note however that a parsimonious solution may look 'simpler' than an intermediate or complex solution, but in fact gives less specific insight. Again an illustration may be helpful.

Assume that Painter A and Painter B are again painting a green painting, but of a much lighter shade than before. Inspecting their paintboxes you find that Painter A has blue, yellow, and white; and Painter B has only blue, and yellow. Based on theoretical knowledge you may assume that adding white to the mix of blue and yellow will result in a light shade of green. Adding this easy counterfactual results in the following

intermediary solution: the configuration of blue, yellow and white paint is sufficient to paint a painting in a light shade of green.

But can you also assume that *not adding* white to the mix will result in a light shade of green? Say that you could (for instance, you could argue that Painter B has painted the painting so thin that the white base layer of the painting shines through, making it overall a lighter shade of green). Using this difficult counterfactual results in the more simplified parsimonious solution: the configuration of blue and yellow paint is sufficient to paint a painting in a light shade of green.

However, the parsimonious solution allows for a much larger set of outcomes than paintings in a light shade of green only. It includes the set of outcomes that includes paintings in any shade of green. This indicates that falsely made assumptions about difficult counterfactuals do however not give 'false' solutions; they just give more inclusive solutions. Though, parsimonious solutions may be 'unrealistically simplistic' (Ragin, 2008, 175).

I have used the expectations (based on the theory on voluntary programmes) as expressed in the article for these minimization steps.

Step 9: Presentation of results

After carrying out the standard analysis results can be presented in various forms. Table 5 in the article is one of the accepted formats for doing so. In the text under table 5 in the article I explain how this table can be read.

Step 10: Interpretation of results (and repeat the steps)

Of course, an fsQCA analysis is but a means to an end and not an end in itself. Once the above analysis has been carried out, the findings should be interpreted in the light of the data obtained. This is what I do in the second half of the article.

Besides, I am not only interested in better understanding the causes of the outcome under scrutiny, but also the causes that have *not* resulted in this outcome. I have carried out that analysis in Section entitled 'Data analysis method: fsQCA', following the above steps. Table A4 provides the truth table for this analysis.

Table A4: Truth table for the analysis of sufficient conditions for outcome ‘not O2’

Row	Conditions							Number (case)	Outcome not O2	Consistency Raw
	<i>Fi</i>	<i>Nm</i>	<i>Le</i>	<i>Pc</i>	<i>En</i>	<i>Gi</i>	<i>GP</i>			
1	1	1	1	1	1	0	0	1	1	1.00
2	1	1	0	1	0	0	0	1	1	1.00
3	1	1	0	0	0	1	0	1	1	1.00
4	1	0	0	1	0	1	1	1	1	1.00
5	1	0	0	0	1	0	0	1	1	1.00
6	1	0	0	0	0	1	0	1	1	1.00
7	0	1	1	1	0	0	1	1	1	1.00
8	1	0	0	1	1	1	0	3	1	1.00
9	1	1	1	1	1	0	1	6	1	1.00
10	1	0	0	0	1	1	0	1	0	0.91
11	1	0	0	0	0	0	0	1	0	0.89
12	1	0	1	1	1	1	0	1	0	0.82
13	0	1	0	1	1	1	1	1	0	0.80
14	1	1	1	1	0	1	0	1	0	0.78
15-128	<i>Logical remainders</i>								?	

Note: The table and symbols are explained below. Abbreviations as per table A2 and the article.

Note, Ragin (2008) advises a consistency score of at least 0.75, which I have followed in the analysis presented under step 7. However, when possible a higher consistency score can be used. It is then advised to look for big jumps in the consistency score column (Schneider & Wagemann, 2012). For this analysis a big jump is identified between row 9 and row 10 (*i.e.* from 1.00 to 0.91). I have therefore chosen a consistency cut off of 1.00 for this analysis.

The findings from the analysis that seeks to understand which configurations of conditions are related to *not* having achieved the outcome is discussed in the article, and the related Table 6 in the article.

References

- Ackerman, F. (1997). *Why do we recycle? Markets, Value and Public Policy*. Washington: Island Press.
- Borck, J., & Coglianese, C. (2009). Voluntary Environmental Programs: Assessing Their Effectiveness. *Annual Review of Environmental Resources*, 34, 305-324.
- Boyd, S. (2013). Financing and Managing Energy Projects Through Revolving Loan Funds. *Sustainability*, 6(6), 345-352.
- Emmenegger, P. (2011). Job security regulations in Western democracies: A fuzzy set analysis. *European Journal of Political Research*, 50(3), 336-364.
- Fiss, P. (2011). Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal*, 52(2), 393-420.
- Fowler, K. M., & Rauch, E. M. (2006). *Sustainable Building Rating Systems: Summary*. Richland: Pacific Northwest National Laboratory.
- Goertz, G., & Mahony, J. (2012). *A Tale of Two Cultures*. Princeton: Princeton University Press.
- Mendel, J., & Korjani, M. (2013). Theoretical aspects of Fuzzy Set Qualitative Comparative Analysis. *Information Sciences*, 237(1), 137-161.
- Policy and Society. (2013). Innovative methods for policy analysis: QCA and fuzzy sets. *Policy and Society*, 32(4), 279-356.
- Ragin, C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

- Ragin, C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: Chicago University Press.
- Rihoux, B. (2013). Qualitative Comparative Analysis (QCA), Anno 2013: Reframing The Comparative Method's Seminal Statements. *Swiss Political Science Review*, 19(2), 233-245.
- Rihoux, B., & Ragin, C. (2009). *Configurational Comparative Analysis*. London: Sage.
- Schneider, C., & Wagemann, C. (2012). *Set-Theoretic Methods for the Social Sciences*. Cambridge: Cambridge University Press.
- Thøgersen, J. (2010). Country Differences in Sustainable Consumption. *Journal of Macromarketing*, 30(2), 171-185.
- Van der Heijden, J. (forthcoming, 2014). *Governance for Urban Sustainability and Resilience: Responding to Climate Change and the Relevance of the Built Environment*. Cheltenham: Edward Elgar.
- Verweij, S., Klijn, E.-H., Edelenbos, J., & Van Buuren, A. (2013). What makes governance networks work? A fuzzy set qualitative analysis of 14 dutch spatial planning projects. *Public Administration*, 81(4), 1035-1055.
- Vis, B. (2012). The Comparative Advantages of fsQCA and Regression Analysis for Moderately Large-N analyses. *Sociological Methods and Research*, 41(1), 168-198.
- Warren, J., Wistow, J., & Bambra, C. (2013). Applying Qualitative Comparative Analysis (QCA) to evaluate a public health policy initiative in the North East of England. *Policy and Society*, 32(4), 289-301.