

# Practical Issues in QCA

Peer C. Fiss  
University of Southern California

*Academy of Management 2009*

Professional Development Workshop on Qualitative  
Comparative Analysis (QCA)

- I. Calibration of fuzzy sets:** how to do it, things to keep in mind, what to avoid
- II. Setting up the analysis:** what is a useful ratio of cases to causal conditions
- III. Interpreting the outcome:** how to read the fsQCA output, what further analysis is required
- IV. Presenting the results:** how can results be shown in a way that is easy to read and interpret

- Creating fuzzy sets using the “direct” method of calibration (Ragin, 2008)
  - The variables are assigned thresholds for full membership, full non-membership, and the crossover point
  - Variables scores are translated into the metric of log odds
  - Membership scores are calculated using the formula below

$$\text{Degree of Membership} = \exp(\log \text{ odds}) / (1 + \exp(\log \text{ odds}))$$

where “exp” stands for the exponentiation of log odds to simple odds

- This method of calibration is now automated in the latest version of the fs/QCA software package

- The laws governing the intersection of fuzzy sets make cases with scores of exactly 0.5 difficult to analyze—fs/QCA will drop them from the analysis
- For these reasons, Ragin (2008) recommends avoiding the use of a precise 0.5 membership score for causal conditions.
- **My approach:** I add a constant of 0.001 to the causal conditions below full membership scores of 1. Adding this constant across all conditions essentially does not affect the results but does assure that no cases are dropped from the fuzzy set analyses

## II. Ratios of Cases to Causal Conditions

- To be able to distinguish “real” from random data, one needs to avoid situations where the ratio of cases to variables drops below a certain threshold
- For very small-N situations (e.g. less than 12 cases), QCA usually adds rather little in terms of additional explanatory power
- Marx (2006) used an experimental design with random data matrices to examine these thresholds. He suggests the following ratios:
  - 4 causal conditions → 10-12+ cases
  - 5 causal conditions → 13-15+ cases
  - 6 causal conditions → 16-25+ cases
  - 7 causal conditions → 27-29+ cases
  - 8 causal conditions → 36-45+ cases

# III. Interpreting Output

- The **Inclusion algorithm** described in Ragin (2000) is the one used in most previous analyses using fuzzy sets
- However, this algorithm circumvents the creation of a truth table and thus forfeits some analytical advantages when e.g. analyzing limited diversity
- To overcome this limitation, Ragin (2005) introduced a **Truth Table algorithm** that is now implemented in the fs/QCA software package (Ragin, Drass, & Davies 2006).
- The algorithm also provides several solutions: **complex**, **parsimonious**, and **intermediate** based on the nature of the simplifying assumptions implemented

\*\*\*\*\*  
 \*TRUTH TABLE ANALYSIS\*  
 \*\*\*\*\*

Model: FSROTA = FSLARGE + FSFORM + FSCENT + FSCMLX + FSDIFF + FSLOWCST + FSMKSPD1 + FSUNCRT

Configs	%	
0 Terms:	10	55.6
1 Terms:	8	44.4
- Terms:	0	0.0
C Terms:	0	0.0
Total:	18	100.0
Dropped:	0	0.0

# Intermediate Solution

\*\*\* TRUTH TABLE SUMMARY \*\*\*

	raw coverage	unique coverage
fslarge*FSCENT*fscmplx*FSDIFF*FSLOWCST*fsmkspd1*fsuncrt+	0.214528	0.012676
fslarge*FSCENT*fscmplx*FSDIFF*fslowcst*fsmkspd1*FSUNCRT+	0.207392	0.035897
fslarge*FSFORM*FSCENT*FSCMLX*FSDIFF*FSLOWCST*fsuncrt+	0.209202	0.028121
fslarge*fsform*fscent*fscmplx*FSDIFF*fslowcst*fsmkspd1*fsuncrt+	0.181401	0.019493
FSLARGE*FSFORM*fscent*FSCMLX*FSDIFF*FSLOWCST*fsmkspd1*fsuncrt	0.150617	0.022582

solution coverage: 0.345759  
 solution consistency: 0.756114

\*\*\* TRUTH TABLE SUMMARY \*\*\*

	Configs	%
0 Terms:	10	55.6
1 Terms:	8	44.4
- Terms:	0	0.0
C Terms:	0	0.0
=====		
Total:	18	100.0
Dropped:	0	0.0

**Parsimonious Solution**

\*\* TRUTH TABLE SOLUTION \*\*

	raw coverage	unique coverage	consistency
	-----	-----	-----
fscmplx*FSDIFF+	0.425861	0.154345	0.740645
FSLOWCST*fsmkspd1+	0.340114	0.034832	0.730664
FSCENT*FSLOWCST	0.338196	0.027375	0.746181

solution coverage: 0.539943

solution consistency: 0.734744

## IV. Presenting Results

- Prior work has faced the challenge of presenting QCA output in a readable form that is easy to interpret
- Too often, the work of interpreting and distilling the results is left to the reader
- This task is even more difficult with the use of the truth table algorithm and the information contained in several solutions (complex, intermediate, parsimonious)
- To remedy this, Ragin & Fiss (2008) have developed a different way of presenting the results in particular of fuzzy set analyses

**Table 7: Configurations for Achieving High Performance**

	Solution					
	1a	1b	2	3a	3b	4
<b>Structure</b>						
Large Size	⊗	⊗	⊗	⊗	⊗	●
Formalization	●	●	⊗	⊗	⊗	●
Centralization	●	●	●	⊗	⊗	⊗
Complexity		●	⊗	●	⊗	●
<b>Strategy</b>						
Differentiation	●	●	●	●	●	●
Low Cost	●	●	●	⊗	⊗	
<b>Environment</b>						
Rate of Change	⊗		⊗	●	⊗	⊗
Uncertainty	⊗	⊗	⊗	⊗	⊗	⊗
Consistency	0.82	0.82	0.86	0.83	0.83	0.82
Raw Coverage	0.22	0.22	0.17	0.14	0.19	0.19
Unique Coverage	0.01	0.01	0.02	0.01	0.02	0.04
<b>Overall Solution Consistency</b>		<b>0.80</b>				
<b>Overall Solution Coverage</b>		<b>0.36</b>				

**Table 8: Configurations for Achieving Very High Performance**

	Solution		
	1a	1b	2
<b>Structure</b>			
Large Size	●	⊗	⊗
Formalization	●	●	⊗
Centralization	⊗	●	●
Complexity	●	●	⊗
<b>Strategy</b>			
Differentiation	●	●	●
Low Cost	●	●	●
<b>Environment</b>			
Rate of Change	⊗	⊗	⊗
Uncertainty	⊗	⊗	●
Consistency	0.83	0.83	0.84
Raw Coverage	0.17	0.22	0.17
Unique Coverage	0.03	0.04	0.03
<b>Overall Solution Consistency</b>		<b>0.81</b>	
<b>Overall Solution Coverage</b>		<b>0.27</b>	

# Empirical Analysis: Configurations and Organizational Performance

- Sample of 205 high-technology manufacturing firms in the UK (Cosh et al., 2002)
- Data collected in 1999 include items on organizational structure, strategy, and environment
- Complete data on performance available for 139 firms; missing values on independent measures imputed using MLE

SECTION B MANAGEMENT ORGANISATION	
IN THIS SECTION WE WOULD LIKE YOU TO TELL US ABOUT YOUR MANAGEMENT ORGANISATION.	
B1. How many levels are there in your business organisation? <i>(Please count the number of levels in the longest line between the direct workers and the chief executive – include both these levels)</i> ..... LEVEL	<input type="text"/>
B2. Which of the following most closely describes the Chief Executive's involvement in decision making? <i>Please tick one box.</i>	INVDEC
Personal control of strategic and operating decisions .....	<input type="checkbox"/> 1
Personal control of strategic decisions, but delegation of operating decisions.....	<input type="checkbox"/> 2
Key member of group taking strategic decisions with delegation of operating decisions .....	<input type="checkbox"/> 3
Other (please specify) .....	<input type="checkbox"/> 4

# Outcome of Interest: **Membership in the Set of High-Performing Firms**

- **Performance** is measured based on Return on Assets (ROA) benchmarked to performance of the high technology sector (median ROA = 7.2%)
- Fuzzy set of *high performing* firms
  - $ROA \geq 16.3\%$  (75<sup>th</sup> percentile) → FS = 1.0
  - $ROA = 11.0\%$  → FS = 0.5
  - $ROA \leq 7.2\%$  (50<sup>th</sup> percentile) → FS = 0.0
- Fuzzy set of *very high performing* firms
  - $ROA \geq 25.0\%$  → FS = 1.0
  - $ROA = 16.3\%$  (75<sup>th</sup> percentile) → FS = 0.5
  - $ROA \leq 7.2\%$  (50<sup>th</sup> percentile) → FS = 0.0

- **Formalization** is measured using a set of 9 survey items that assess to what extent e.g.
  - Formal policies and procedures guide decisions
  - Communications are documented by memos
  - Reporting relationships are formally defined
  - Plans are formal and written
- Items combined into a scale (Cronbach's  $\alpha = .83$ )
- Fuzzy set of firms with high degree of *formalization*
  - "Nearly always" → FS = 1.0
  - "About half the time" → FS = 0.5
  - "Almost never" → FS = 0.0

- **Centralization** is measured using a set of 5 survey items that assess who is the last person whose permission must be obtained (“department head, division head, CEO, Board of Directors”) for e.g.
  - Addition of a new product or service
  - Unbudgeted expenses
  - Selection of type or brand of new equipment
- Items combined into a scale ( $\alpha = .74$ )
- Fuzzy set of firms with high degree of *centralization*

“Board of Directors”	→	FS = 1.0
<i>scale mid-point</i>	→	FS = 0.5
“Department Head”	→	FS = 0.0

- **Complexity** is measured using a combined measure of vertical and horizontal differentiation
  - Vertical differentiation was measured as the number of levels in the longest line between direct worker and CEO (Pugh et al., 1968)
  - Horizontal differentiation was measured using the number of functions with at least one full-time employee (Pugh et al., 1968)
  - Complexity is calculated as the product of horizontal and vertical differentiation (Singh, 1986; Wong & Birnbaum More, 1994)
- Fuzzy set of firms with high degree of *complexity*
  - 99<sup>th</sup> percentile (6 Levels / 17 Functions) → FS = 1.0
  - 50<sup>th</sup> percentile (3 Levels / 9 Functions) → FS = 0.5
  - 1<sup>st</sup> percentile (1 Level / 1 Function) → FS = 0.0

- **Size** is measured as average number of full time employees, with fuzzy set membership tied to US SME categories
- Fuzzy set of *large* firms
  - 250+ employees → FS = 1.0
  - 50 employees → FS = 0.5
  - 10 or less employees → FS = 0.0

- **Differentiation strategy** measured as competitive capability based on product features and new product introduction, combined into 5-point scale ( $\alpha = .80$ )
- **Low cost strategy** measured as competitive capability based on low labor cost, material cost, energy consumption, inventory cost, combined into 5-point scale ( $\alpha = .86$ )
- Recoded into two fuzzy sets of firms with a *differentiation strategy* and *low cost strategy*
  - (5) "Critically important" → FS = 1.0
  - (3) *scale mid-point* → FS = 0.5
  - (1) "Not important" → FS = 0.0

- **Rate of Change** assesses how fast the environment is changing and is measured as length of main product life cycle in months, recoded into fuzzy set of firms operating in a *high velocity environment*
  - 1 months → FS = 1.0
  - 36 months → FS = 0.5
  - 120 months → FS = 0.0
- **Uncertainty** is measured using two items that assess how predictable were technological changes in the environment, combined into a scale ( $\alpha = .74$ ) and recoded into a fuzzy set of firms operating in a *highly uncertain environment*
  - "Completely unpredictable" → FS = 1.0
  - scale mid-point* → FS = 0.5
  - "Easily predictable" → FS = 0.0

- All variables were transformed into fuzzy sets using the “direct” method of calibration (Ragin, 2008)
  - The variables are assigned thresholds for full membership, full non-membership, and the crossover point
  - Variables scores are translated into the metric of log odds
  - Membership scores are calculated using the formula below

$$\text{Degree of Membership} = \exp(\log \text{ odds}) / (1 + \exp(\log \text{ odds}))$$

where “exp” stands for the exponentiation of log odds to simple odds

- The rescaled measures range from 0 to 1 and are tied to their respective membership thresholds and crossover points

- Clustering on four structure and two strategy variables
- Two-step cluster analysis (Ketchen and Shook, 1996;
  - hierarchical cluster analyses using Ward's minimum variance method suggested a three-cluster solution based on cutoff values
  - K-means cluster analysis with centroid values of hierarchical analysis as seeds (e.g. Payne 2006; Lim et al., 2006)
- Results are essentially stable across different clustering algorithms
- Tobit regression of dummy variables for cluster membership on performance measure with interaction terms for environmental rate of change and uncertainty

# Analysis I: Cluster Analysis

	Final Cluster Centers		
	"Prospector"	"Analyzer"	"Defender"
<b>Structure</b>			
Size	0.13	0.22	0.64
Formalization	0.52	0.49	0.76
Centralization	0.62	0.52	0.39
Complexity	0.40	0.42	0.71
<b>Strategy</b>			
Differentiation	0.82	0.19	0.75
Cost Leadership	0.37	0.42	0.42
<b>N</b>	93	41	59

Table 5: T

Independent

Prospector	1.137* (0.532)	1.159† (0.700)	0.264 (0.771)	1.123* (0.528)	1.154* (0.527)
Defender	0.130 (0.521)	0.129 (0.522)	0.081 (0.515)	0.796 (0.721)	0.933 (0.807)
Environmental Pace	1.146 (0.710)	1.175 (0.918)	1.029 (0.697)	1.817* (0.903)	1.158 (0.705)
Environmental Uncertainty	-0.853 (0.700)	-0.850 (0.704)	-1.649 (0.920)	-0.830 (0.692)	-0.206 (0.829)
Prospector × Rate of Change		-0.068 (1.383)			
Prospector × Uncertainty			1.977 (1.406)		
Defender × Rate of Change				-1.937 (1.448)	
Defender × Uncertainty					-1.910 (1.466)
Constant	0.793 (0.542)	0.783 (0.583)	1.196 (0.612)	0.568 (0.560)	0.488 (0.579)
LR Chi-squared	11.66*	11.66*	13.72*	13.53*	13.43*
Pseudo R-Squared	0.04	0.04	0.05	0.05	0.04

<sup>a</sup> Standard errors in parentheses.

†  $p \leq .10$ ; \*  $p \leq .05$ ; \*\*  $p \leq .01$ ; \*\*\*  $p \leq .001$

# Analysis II: Deviation Scores

- Profiles are defined based on the ideal types of prospector, analyzer, and defender (Miles & Snow, 1978)
- For each profile,  $D_{io}$  is calculated as the deviation of organization  $o$  from ideal type  $i$  and across all attributes
- Ideal profile fit is calculated as the minimum deviation across the three profiles (Doty, Glick, & Huber, 1993; Doty & Glick, 1994)

$$Fit_{IT} = - \left( \min_{i=1}^l D_{io} \right)$$

- Tobit regression of minimum and individual fit measures on performance with interaction terms for environmental velocity and uncertainty

# Analysis II: Deviation Scores

	Ideal Profiles Based on Miles & Snow (1978)		
	Prospector	Analyzer	Defender
<b>Structure</b>			
Size	Low	Medium	High
Formalization	Low	Medium	High
Centralization	Low	Medium	High
Complexity	Low	High	High
<b>Strategy</b>			
Differentiation	High	Medium	Low
Low Cost	Low	Medium	High

Ideal Type Fit (min)	0.801† (0.598)							
Prospector Dev.	-0.212 (0.280)				0.082 (0.422)	0.814 (0.529)		
Analyzer Dev.			-0.023 (0.456)				-0.509 (0.677)	-0.485 (0.791)
Defender Dev.					0.212 (0.280)			
Prospector Dev. × Rate of Change					-0.800 (0.886)			
Prospector Dev. × Uncertainty							-2.077* (0.951)	
Analyzer Dev. × Rate of Change							1.584 (1.649)	
Analyzer Dev. × Uncertainty							1.049 (1.470)	
Environmental Rate of Change	1.186 (0.718)	1.227† (0.731)	1.161 (0.725)	1.227† (0.731)	3.444 (2.608)	1.140 (0.713)	-1.638 (2.954)	1.162 (0.722)
Environmental Uncertainty	-0.989 (0.714)	-1.012 (0.728)	-0.925 (0.724)	-1.012 (0.728)	-1.001 (0.723)	4.708† (2.638)	-0.919 (0.718)	-2.814 (2.767)
Constant	2.725* (1.130)	1.959* (0.901)	1.415 (0.901)	0.686 (0.989)	1.159 (1.227)	-0.911 (1.488)	2.275† (1.281)	2.228 (1.467)
LR Chi-squared	6.21*	4.91	4.33	4.91	5.75	10.61*	5.27	4.85
Pseudo R-squared	0.02	0.02	0.03	0.02	0.02	0.04	0.02	0.02

<sup>a</sup> Standard errors in parentheses. T-tests are one-tailed where predicted, otherwise two-tailed.

† p ≤ .10; \* p ≤ .05; \*\* p ≤ .01; \*\*\* p ≤ .001

- The **Inclusion algorithm** described in Ragin (2000) is the one used in most previous analyses using fuzzy sets
- However, this algorithm circumvents the creation of a truth table and thus forfeits some analytical advantages when e.g. analyzing limited diversity
- To overcome this limitation, Ragin (2005) introduced a **Truth Table algorithm** that is now implemented in the fs/QCA software package (Ragin, Drass, & Davies 2006). This algorithm additionally allows the calculation of consistency and coverage scores

- Set-theoretic consistency assesses the degree to which cases sharing a given condition or combination of conditions agree in displaying the outcome in question
- With **crisp sets**, a straightforward measure of consistency is the proportion of cases with a given cause (or combination of causes) that also exhibit the outcome
- With **fuzzy sets**, the case is somewhat more complex. When perfectly consistent, a fuzzy subset relationship is characterized by a plot with a triangular shape

# Evaluating Consistency

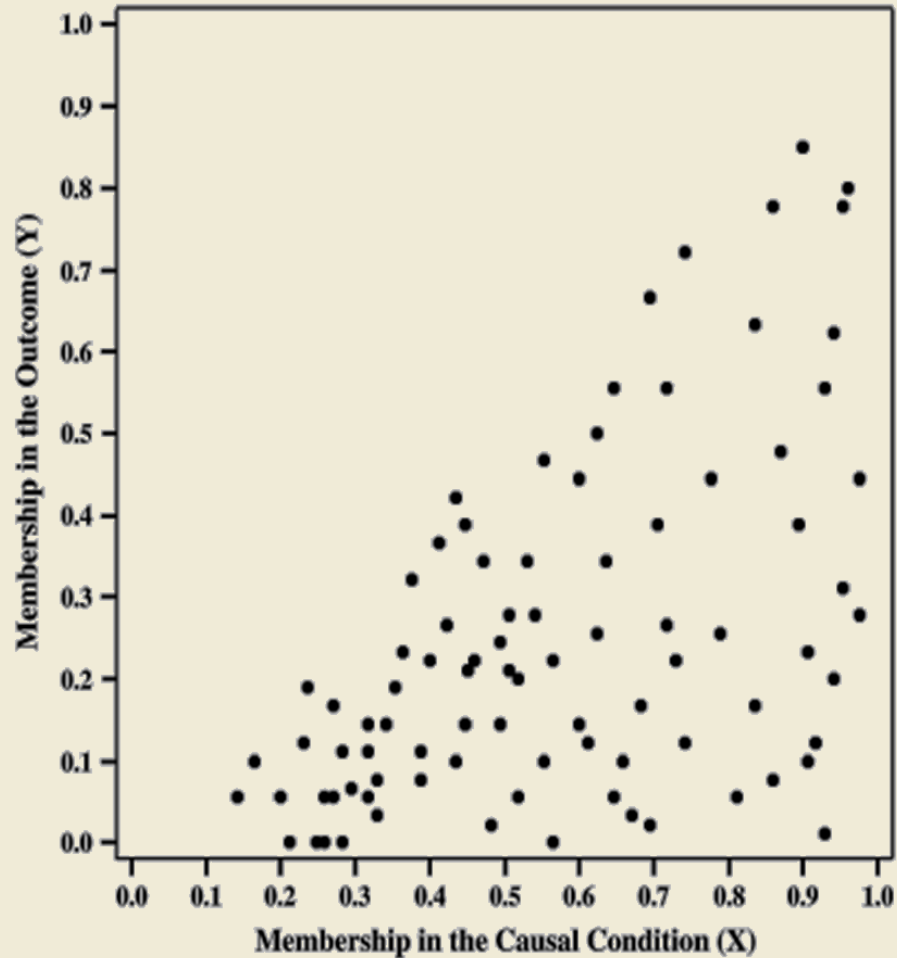
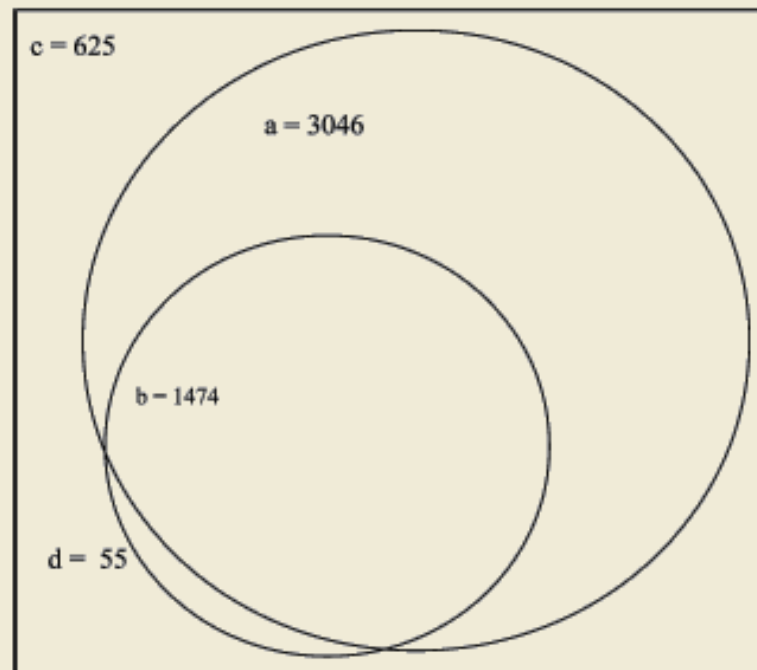


Fig. 2 Fuzzy subset relation consistent with necessity.

# Evaluating Consistency

- A simple way to characterize consistency of fuzzy sets is the proportion of cases consistent with the outcome. However, this gives undue weight to cases displaying only low membership in the condition of interest
  - For example, a case with **high membership** in the outcome should have **more weight** than one with **low membership**
- An alternative measure: the sum of **consistent** fuzzy membership scores divided by the sum of **all** membership scores
- This measure is further refined by adding to the numerator the part of each inconsistent causal membership score that is consistent with the outcome. This adjustment gives **credit for near misses and penalties for large inconsistencies**
- For the current analyses, minimum acceptable consistency for the solutions was set at 0.80 (Ragin, 2008)

- Coverage is a measure of how important a cause or causal combination is to the outcome. It is similar to an R-Square by indicating the number of cases that take this path to the outcome



- Coverage is useful because it allows the researcher to evaluate the importance of different causal paths. For example, conditions may be necessary but empirically irrelevant due to low coverage (e.g. only very few cases in a large dataset). Again, this is similar to practical significance in an R-Squared measure
- Coverage is computed only after consistency has been established and there can be tradeoffs between both (e.g. a very consistent combination may have low coverage and a combination with high coverage may have low consistency)
- Coverage can be further subdivided into unique coverage of causal conditions (similar to unique R-Squared calculations in regression analysis). Coverage is also incorporated into the latest version of fs/QCA

**Table 7: Configurations for Achieving High Performance**

	Solution					
	1a	1b	2	3a	3b	4
<b>Structure</b>						
Large Size	⊗	⊗	⊗	⊗	⊗	●
Formalization	●	●	⊗	⊗	⊗	●
Centralization	●	●	●	⊗	⊗	⊗
Complexity		●	⊗	●	⊗	●
<b>Strategy</b>						
Differentiation	●	●	●	●	●	●
Low Cost	●	●	●	⊗	⊗	
<b>Environment</b>						
Rate of Change	⊗		⊗	●	⊗	⊗
Uncertainty	⊗	⊗	⊗	⊗	⊗	⊗
Consistency	0.82	0.82	0.86	0.83	0.83	0.82
Raw Coverage	0.22	0.22	0.17	0.14	0.19	0.19
Unique Coverage	0.01	0.01	0.02	0.01	0.02	0.04
<b>Overall Solution Consistency</b>		<b>0.80</b>				
<b>Overall Solution Coverage</b>		<b>0.36</b>				

**Table 8: Configurations for Achieving Very High Performance**

	Solution		
	1a	1b	2
<b>Structure</b>			
Large Size	●	⊗	⊗
Formalization	●	●	⊗
Centralization	⊗	●	●
Complexity	●	●	⊗
<b>Strategy</b>			
Differentiation	●	●	●
Low Cost	●	●	●
<b>Environment</b>			
Rate of Change	⊗	⊗	⊗
Uncertainty	⊗	⊗	●
Consistency	0.83	0.83	0.84
Raw Coverage	0.17	0.22	0.17
Unique Coverage	0.03	0.04	0.03
<b>Overall Solution Consistency</b>		<b>0.81</b>	
<b>Overall Solution Coverage</b>		<b>0.27</b>	

- In fuzzy set analysis, an important aspect relates to modeling the absence of the outcome
  - In this case that means modeling the absence of high performance; note that this is **different** from modeling causes leading to low performance
- Using the negation of the outcome here leads to consistency scores considerably below the acceptable level of 0.75, indicating the absence of a clear set-theoretic relationship
- **Put differently**: there are few configurations that consistently lead to high performance, but no consistent path to average performance high performance
- QCA thus allows for the analysis of **causal asymmetry**, a concept foreign to correlational methods that always conceive of causal relations in symmetric terms

## What we can learn from others...

- Be cohesive; realize that we are still a growing community and need to work with each other
- Know and acknowledge each other's work
- What is most needed—the role of A-journal publications
- Overcoming the current double hurdle of set-theoretic research
- Institutionalizing the approach