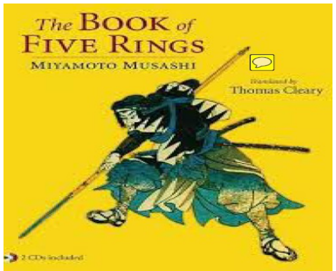
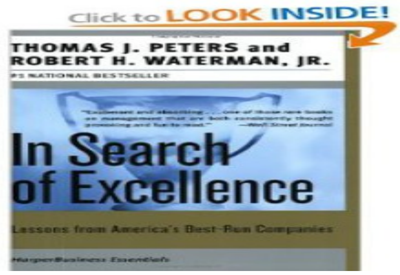
Lecture 19: QCA & Boolean Truth Tables: the "right way" to see case studies

Business schools use case studies – detailed analysis of a particular situation – to teach about decision-making. Top-selling business books generalize from a limited number of cases. In Search of Excellence, published in 1982 and one of the most widely read business books ever, selling 3 million copies in its first four years, was based on a few successful 1970s companies (some of which later bombed) **with no examples of failures to contrast with!** ***The Book of Five Rings*** (五輪書 *Go Rin No Sho*) by the swordsman Miyamoto Musashi circa 1645 was sold as a great business strategy book in the 1980s when American business believed that Japan had the right answers to business success (Ezra Vogel, Japan as Number One).





QCA — qualitative comparative analysis — is a way to analyze problems where there are few observations that highlights the importance of **CONFIGURATIONS** of variables on outcomes. It uses **Boolean truth tables** to provide a logical framework for B-School CASE studies. http://www.compasss.org/ is the portal web site for QCA stuff. It lists programs that you can freely download such as http://www.u.arizona.edu/~cragin/fsQCA/software.shtmlI

It uses fuzzy set theory i FUZZY QCA in place of Boolean crisp set models. http://www.stata-journal.com/article.html?article=st0140: " This article briefly discusses the substantive motivation and technical details of QCA, as well as fuzzy-set QCA , followed by … discussion of how the new program fuzzy performs these techniques in Stata. An empirical example is presented that demonstrates the full suite of tools contained within fuzzy, including creating configurations, performing ... statistical tests of the configurations, and reducing the identified configurations."

STATA:Fuzzy command... is capable of creating, testing, and performing logical reductions on both fuzzy and dichotomous (crisp) set-theoretic data. It has a capacity to **test sets of configurations for logical necessity and sufficiency probabilisticly.** The user has to specify which variate is the outcome. Fuzzy also has the ability to perform other useful statistical tests. ... For new users to the fuzzy command ...type: find it fuzzy in command box.

I. Cases as Well-Specified Configuration

A case study teachers you a lot about a particular situation. The key question is whether you can generalize to other situations. Or, if you have a few cases, how many you need to do more than tell a "war story"/anecdote? An anecdote — a single case — is better than nothing but easy to "over-fit". as you can often find a consistent explanation for whatever happened in that case but have no way to test your explanation. You need a "counter-factual" to your interpretation of the case – say that X caused success – that could disprove you. Success without X? A failure with X?

QCA is a tool for answering these questions that exploits the notion of a **configuration** — a specific set of attributes that explain a phenomenon. Recall landscape models. If the landscape of configurations is rugged, it is hard to generalize. With lots of configurations and few cases, you are limited and at risk of over-fitting/over-generalizing.

Linear regression models assume that variables are related by a **linear** form that has parsimony because it interpolates/extrapolates using few parameters: **P =** a +bX +cY + dZ + e — where the 4 parameters are a, b, c, d and e is an error. With 4 parameters and e you need 5 or more data points to fit. Note that dP/dX = b does not depend on the Y or Z.

But if configurations of inputs matter, interpolating/extrapolating linearly is not advisable. Ragin uses term MULTIPLE CONJUNCTURAL CAUSATION for situations in where configurations matter. Consider a full linear interaction model where there are linear terms and multiplicative two variable terms, and three variable terms

$$P = a + bX + cY + dZ + fXY + gYZ + hXZ + i\,XYZ + e.$$

Here $dP/dX = b + fY + hZ + iYZ$ so the impact of X depends on the values of Y and Z. (Recall Kauffman's N-K landscape model, where K is the number of factors that determine the profit from a given value of one factor). In a 3 variable case, where each variable is (0/1), you need $2^3 = 8$ observations to cover all possible cases. If the variables take more values, you need more observations. Configurations --> # of observations rises rapidly even in the 0/1 case

# of variables	# of observations needed	
	Regression	Configuration
3	5	8

5	7	32
7	9	128
10	12	1024

In a CASE STUDY you see only a few configurations. But you want to make a general inference.

Example: students from some universities do better than others in knowledge of folk music. You want to how the following five attributes of a university produce this result: admission policy; sterling faculty; urban location; Hillbilly music on Sat Morning; diversity of student body. Using a 0/1 code for the attributes, you have 32 configurations. But your data only covers Yale, BU, Princeton, Dartmouth and Harvard. What can you infer from 5 of 32 cases?

Case-Oriented Approach: road to understanding social phenomena is by closely examining individual cases. Measured variables incompletely represent phenomena. Best way to "fill in critical details" is through close engagement and intensive examination. LONG INTERVIEW FOR JOB. History. Law. Medicine. Big data along the variable dimension instead of the number of observation dimension.

Variable-Oriented Statistical Approach – find broad patterns across many cases Individual cases too idiosyncratic to generalize. Statistical averages reveal underlying structures and relationship.

As in the Genetic Algorithm, QCA uses * to deal with don't knows or missing observations. With few cases, best to view each conjunction as **accurate.** Fuzzy QCA modifies this, but the accuracy assumption makes it easy to see how QCA deals with the problem via **Boolean Algebra**

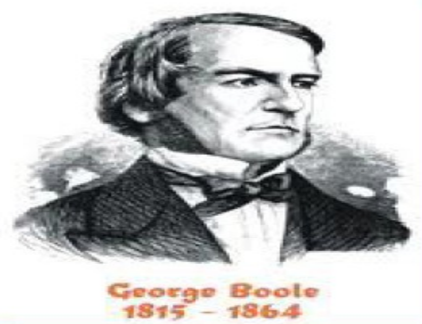


TABLE 1-1 Basic Identities of Boolean Algebra

(1) $x + 0 = x$	(2) $x \cdot 0 = 0$
(3) $x + 1 = 1$	(4) $x \cdot 1 = x$
(5) $x + x = x$	(6) $x \cdot x = x$
(7) $x + x' = 1$	(8) $x \cdot x' = 0$
(9) $x + y = y + x$	(10) $xy = yx$
(11) $x + (y + z) = (x + y) + z$	(12) $x(yz) = (xy)z$
(13) $x(y + z) = xy + xz$	(14) $x + yz = (x + y)(x + z)$
(15) $(x + y)' = x'y'$	(16) $(xy)' = x' + y'$
(17) $(x')' = x$	

II.Boolean Algebra and Truth Tables

The 4 steps of QCA:
1. Describe observations as a Boolean Truth Table, caps for presence, little letters for absence
2. Combine statements via Boolean minimization — mechanical operation
3. Make assumptions about missing observations to draw generalizations
4. Build more detailed models to deal with aberrant cases

1.Boolean algebra:

OR: A+B+C means A or B or C is present

AND: ABC means A and B and C are all present

NOT A= A' where use a to be not A

So Y = ABC means that when we have ABC we get Y; Y = A+B+C means that if we have A or B or C we get Y
Boolean Algebra lives on De Morgan's laws:

(AB)' = a + b : not having A and B means you must have a or b
(A+B)' = ab: not having A or B means you must have ab
A' = a (if A=1, a=0; if A=0, a=1)

Now consider a truth table. Here is one with 16 cases (Ragin, p 89):

Representative Truth Table with Four Causal Conditions

A	B	C	D	Y	# Instances
0	0	0	0	0	8
0	0	0	1	0	6
0	0	1	0	1	10
0	0	1	1	0	5
0	1	0	0	1	13
0	1	0	1	0	7

0	1	1	0	1	11
0	1	1	1	1	5
1	0	0	0	1	9
1	0	0	1	1	3
1	0	1	0	0	12
1	0	1	1	0	23
1	1	0	0	0	15
1	1	0	1	1	5
1	1	1	0	0	8
1	1	1	1	1	6

Write the table as the Boolean sum of observations

Y = abCd + aBcd + aBCd ... (ie the third line + the fifth line + ... all lines where Y is1

y = Y' = abcd + abcD etc (all lines where Y is 0).

This is a COMPLETE DESCRIPTION of the data, but not easy to interpret, so

2)M**ake a simpler statement via Boolean Minimization**

The y statement above can be simplified since **abcd+abcD = abc,** which says that **i**t doesn't matter what D is: abc → y . Also, AbCD+ ABCD = ACD. If you have Y (or y) in cases where ACD is present when B is present and absent, surely B does not matter.

Use **prime implicants** – terms that are like prime numbers in that they cannot be derived from other terms – to get parsimonious logical structure. QCA programs do this for you. You can simplify in several ways. Perhaps you want most parsimonious. (If you have a theory/concern over X, you will want to test your theory).

Prime Implicant Chart. The primitives are given in the columns. The primes are in the rows. The xs show what you can simplify. ABc and aBc simplify to Bc. You can use the same terms over and over in the the simplification.

	ABC	AbC	ABc	aBc
AC	x	x		
AB	x		x	
Bc			x	x

In the first row AC simplifies ABC and AbC – B does not matter.

The goal is to generalize from the few cases to a broader statement that is CONSISTENT with the data. We do this by making assumptions.

3) BUT truth table with information on all 16 cases is unrealistic. You rarely observe all configurations. Say you had information on 8 cases, with no observations for the others. What might you do?

Write out Boolean statement for 0 (observed) and o (not observed). Perhaps there is a reason you do not observe some cases: maybe they are "impossible" — ie no school has bad admission policy and sterling faculty (faculty would leave) or selective admission and bad faculty (students would leave).

o = abcD + abCd + ... ; 0 = abcd + aBcd +

If missing cannot exist, this is perfect. When you *assume* missing cannot occur, you are making a stronger theory.

b). **POSTULATE what the missing cases might show--** *widening your theory*.

Postulate that all of the missing are 0 or 1. This will give you a simpler expression.

If you set missing observations to 0, your Boolean statement gives **ONLY the conditions** under which you get Y. You assume that under all other conditions you do not get Y. This is a conservative statement of what configurations you know produce Y.

If you set missing observations to 1, you specify the conditions under which it is possible for Y to exist — you have no information that Y cannot exist under 1, so you report the **possible conditions for Y.** When you assume the missing observations are 1 — you are stating possible event.

4.Correcting Errors: What do you do about the PY' or P'Y cases, where your prediction failed?

Ragin that you should study contradictions carefully -- find out what is wrong -- and modify your theory. This can involve changing your theory or going back and getting new data -- you think the theory fails because of an additional factor, which you never measured. So measure it. But it risks **over-fitting.**

Example: You investigate productivity You have 20 firms; of whom 10 are highly productive and have profit-sharing, teamwork, training, and unionization. Your initial hypothesis is that Profit-Sharing —> Productivity. You find

Profit-sharing	Teamwork	Training	Union	Outcome
1				6 of 10
0				4 of 10
NOW ADD TEAMWORK: NOTE IF START WITH TEAMWORK 7 of 11 good vs 4 of 9 not good				
1	1			4 of 6
1	0			2 of 4
0	1			3 of 5
0	0			2 of 5
NOW ADD TRAINING: By itself training would be 8 of 12				
1	1	1		4 of 4
1	1	0		2 of 4
1	0	1		0 of 2
1	0	0		
0	1	1		3 of 5
0	1	0		1 of 1
0	0	1		
0	0	0		0 of 4

But with lots of cases, CERTAIN to get some contradictions. So it might be sensible be to determine a "plausible" level of error that you would regard as consistent with your model. Looking at data you will note that when you have Teamwork and Training — 7/9 77% productive; Other 3/11 27% That is a pretty good result.

You could also revise your theory to a fuzzy one by changing your theory from Teamwork and training—> Productivity to Teamwork and training produce high productivity "most of the time".

QCA forces you to think about CONFIGURATIONS and "dialogue" with data and clarify your views about non-existent situations. But danger of over-fitting cases --> need for A TEST SET to see if your reading of case study/QCA generalizes.

Example 1– "Limits to non-state market regulation:A qualitative comparative analysis of the international sport footwear industry and the **Fair Labor Association"** A. Marx, (Regulation & Governance (2008) 2, 253-273)

The FLA (www.fairlabor.org/fla) and is a multi-stakeholder coalition of industrial companies, colleges and universities, and NGOs founded in 1997 to protect workers worldwide and provide public information to consumers …FLA developed a code of conduct on forced labor, child labor, harassment or abuse, discrimination, health and safety, freedom of association and collective bargaining, wages and benefits, hours of work, and overtime compensation. Several apparel and footwear firms sourcing from over 3,000 suppliers in 80 countries joined and committed to implement FLA code. Often works with rival student run Workers Rights Consortium www.workersrights.org/

FLA compliance regime: 1) companies have a compliance program, designating someone responsible for promoting code compliance throughout supply chain; 2) an internal system of monitoring factory conditions; 3) allowing and facilitating independent external monitoring; 4) correcting problems at a non-compliant factory rather than closing it and moving operations elsewhere; 5) annual audits of compliance program to review procedures and documentation and field visits to observe the work of local compliance staff and assess factory conditions; 6) public disclosure. .

Marx examines 17 international sport footwear companies and looks for why six (**Nike, Adidas, Puma, Reebok, Asics and Umbroj)** joined FLA. Finds that combination of sustained NGO pressure and public ownership of a firm is a necessary precondition for firms joining a multi-stakeholder certification initiative. One additional variable: Prior Change --Companies which had a code and implementation mechanism prior to joining certification coded 1. Companies which had a code but no implementation mechanism are coded 0.

Table 1 NGO pressure on firms – identification and QCA scores

Firm	Pressure by NGOs	QCA score
Nike	Sustained pressure	1
Adidas	Sustained pressure	1
Puma	Sustained pressure	1
Reebok	Sustained pressure	1
New Balance	Sustained pressure	1
Diadora	Some negative reports by local NGOs but not sufficient to be regarded as sustained pressure	0
Fila	Sustained pressure	1
Karhu	No sustained pressure	0
Kelme	No sustained pressure	0
Mizuno	No sustained pressure	0
Saucony	No sustained pressure	0
Asics	Sustained pressure from Oxfam 2004 and Dutch Advertising code commission	1
Brooks	No sustained pressure	0
Decathlon	No sustained pressure	0
Lotto	No sustained pressure	0
Kappa	No sustained pressure	0
Umbro	Sustained pressure in the context of the London bid for Olympic Games 2012	1

Table 7 Data matrix

Firm	Home UNIONS	Sustained NGO pressure	Public	Prior change	Member FLA
Nike	0	1	1	1	1
Adidas	1	1	1	1	1
Puma	1	1	1	1	1
Reebok	0	1	1	1	1
New balance	0	1	0	1	0
Diadora	1	0	0	0	0
Fila	1	1	1	0	0
Karhu	1	0	0	0	0
Kelme	–	0	0	0	0
Mizuno	1	0	1	0	0
Saucony	0	0	1	0	0
Asics	1	1	1	1	1
Brooks	0	0	0	0	0
Decathlon	1	0	0	1	0
Lotto	1	0	0	0	0
Kappa	1	0	1	1	0
Umbro	0	1	1	0	1

This produces 16 possible configurations but in fact only 12 found in world (* is the AND; + is the OR)

The conditions under which firms join can be represented by NGO*PUBLIC*CHANGE + union*NGO*PUBLIC. NGO AND PUBLIC are the necessary preconditions (but not sufficient. Why?)

The conditions under which firms do not join the FLA does not so easily simplify:
ngo*change + UNION*ngo + UNION*PUBLIC*change + union*NGO*public*CHA

Table 8 Overview of configurations explaining joining or not joining FLA (listing truth table)

	OUTCOME = 1 (Joining FLA)	N
1	UNION*NGO*PUBLIC*CHANGE	3
2	union*NGO*PUBLIC*CHANGE	2
3	union*NGO*PUBLIC*change	1
	OUTCOME = 0 (Not joining FLA)	
4	UNION*ngo*public*change	3
5	union*ngo*public*change	1
6	union*ngo*PUBLIC*change	1
7	UNION*ngo*PUBLIC*change	1
8	UNION*ngo*public*CHANGE	1
9	union*NGO*public*CHANGE	1
10	UNION*ngo*PUBLIC*CHANGE	1
11	UNION*NGO*PUBLIC*change	1
12	ngo*public*change	1
	Total	17

Practical problem with QCA is that you often do not get nice simplifications. What to do next?
Go Fuzzy; Add more variables; Say world is complex and ...

Example 2: R Dekker and P Scholten "Framing the Immigration Policy Agenda: A Qualitative Comparative Analysis of Media Effects on Dutch Immigration Policies" The International Journal of Press/Politics 2017, Vol. 22(2)

Studies sixteen focusing events related to Dutch immigration policies to determine conditions where media coverage was associated with changes on the policy agenda. Differentiates four frames: **human-interes**t which portrays immigrants and refugees as victims who require compassion and help; **threat frame** which presents immigrants as threat to host society; **economic** frame: **managerialis**t frame which stresses governance challenge. Finds that *Quantity of media attention and frame consonance are relevant indicators of changes on the policy agenda only when the majority of media coverage is contesting the current policy frame.*"

Table 1. Operationalization of Conditions for Analysis.

Outcome: Policy change	Change in issue frame on the policy agenda within a period of maximum one year after initiation of media coverage of the case.
Media attention	More than 100 publications in our selection of newspapers, opinion magazines, and television programs within six months after the onset of media coverage of the case.
Frame contestation	When the framing of more than 50% of media coverage is different from the initial issue frame on the policy agenda.
Frame consonance	When one frame is dominant in more than 50% of all media publications.

Table 2. Data Matrix "Policy Change" and Three Conditions (Crisp Set).

Case ID	Media Attention (1 = Large)	Frame Contestation (1 = Yes)	Frame Consonance (1 = Yes)	Policy Change (1 = Yes)
MAN	1	1	0	1
DOL	1	1	0	1
AHM	0	1	1	1
BUT	0	1	1	1
ORA	0	1	0	1
PAR	0	1	0	1
AUP	0	1	1	1
ERI	1	0	0	0
UGA	0	0	0	0
RET	0	0	1	0
AMS	1	1	0	0
IJS	0	0	0	0
HUN	0	1	1	0
CHI	0	1	1	0
POL	1	1	0	0
COD	0	0	1	0

Note. AMS = case of Vluchtkerk; AHM = case of Abdul Ghafoor Ahmadzai; AUP = new au pair policy; BUT = case of Dennis Butera; CHI = reunification of children; COD = official warning about labor mobility; DOL = case of Aleksandr Dolmatov; ERI = Eritrean asylum seekers; HUN = hunger strike; IJS = asylum center at IJsselhallen; MAN = case of Mauro Manuel; ORA = asylum center in Oranje; PAR = amnesty for asylum children; POL = case of Polenmeldpunt; RET = return of rejected asylum seekers; UGA = Ugandan gay immigrants.

Table 3. Truth Table Policy Change According to Conditions Media Attention, Frame Contestation, and Frame Consonance (C = Contradictory Row; R = Logical Remainder).

Media Attention (MA)	Frame Contestation (FT)	Frame Consonance (FS)	Policy Change (PC)	Cases
0	1	0	**1**	ORA; PAR
1	0	0	**0**	ERI
0	0	0	**0**	UGA; IJS
0	0	1	**0**	RET; COD
1	1	0	**C**	MAN; DOL; AMS; POL
0	1	1	**C**	AHM; BUT; CHI; AUP; HUN
1	1	1	**R**	—
1	0	1	**R**	—

Table 4. QCA Minimization Results.

Outcome (PC)	Configuration	*N*
C	MA × FT × fs + ma × FT × FS	9
1	ma × FT × fs	2
0	ft × fs + ma × ft	5

Note. QCA = qualitative comparative analysis; PC = policy change; MA = media attention; FT = frame contestation; FS = frame consonance. Lower case = absence of condition; × = AND; + = OR.

YOUTUBE VIDEOS: Wendy Olsen https://www.youtube.com/watch?v=q917kpR9iGM
https://www.youtube.com/watch?v=2TxtjMDpDgM