

## The distinctiveness of case-oriented research:

In a recent article in *Studies in Comparative International Development*, Christopher Achen, a well-known quantitative researcher, notes:

Contemporary case-study methods are difficult to explicate in conventional statistical theory, and yet they are frequently quite powerful and successful in ways that no statistical methods could match. *An important clue is that they often carry out an implicit comparison against known background relationships*, most obviously so in single-case studies (Ragin 2000: 206). But what is the precise inferential logic of this step and why is it so successful? *No one knows.* (italics added)

## Limited diversity:

Naturally occurring social phenomena (i.e., nonexperimental data) are almost always limited in their diversity.

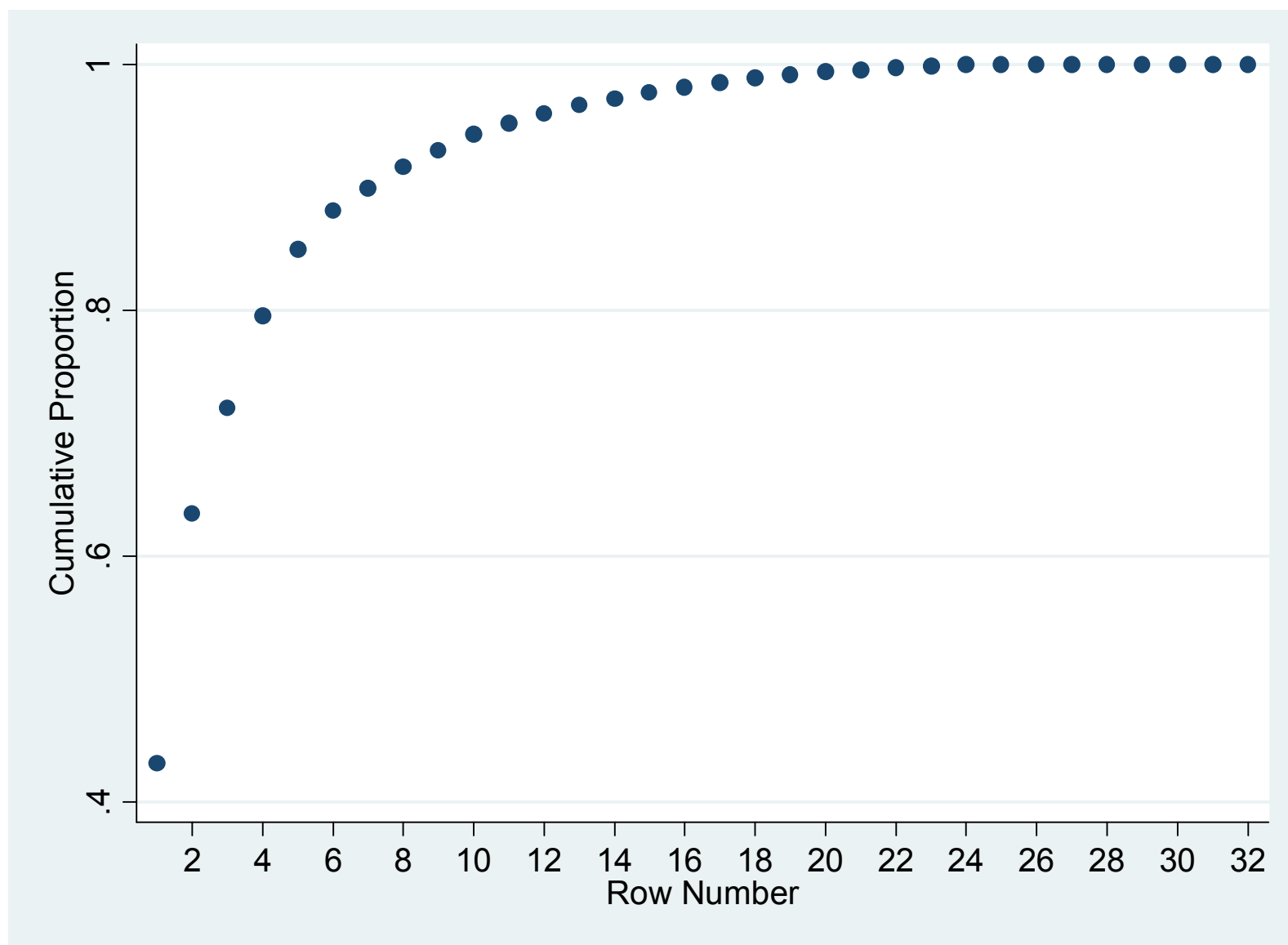
Limited diversity is easy to visualize. Consider a multi-dimensional vector space defined by five causal conditions or “independent variables.” There are 32 sectors ( $2^5$ ) in this vector space.

In the typical social science data set many of these sectors are void or virtually void of cases. Sectors without cases can be seen as potential counterfactual cases.

While many social scientists think that limited diversity is a small-N problem, it is common in large-N analyses as well, but rarely recognized.

# Limited diversity in a typical large-N social science data set:

College	High parental inc.	High AFQT	Married	Children	Freq.	Cum Freq	Cum prop.
0	0	0	0	0	327	327	0.431398
0	0	0	0	1	154	481	0.634565
0	0	0	0	0	65	546	0.720317
1	0	0	0	0	57	603	0.795515
0	0	0	0	1	41	644	0.849604
1	0	0	0	1	24	668	0.881266
0	1	0	0	0	14	682	0.899736
1	0	0	0	1	13	695	0.916887
0	1	0	0	1	10	705	0.930079
1	1	0	0	0	10	715	0.943272
1	0	0	0	1	7	722	0.952507
1	0	1	1	1	6	728	0.960422
0	1	0	0	1	5	733	0.967018
1	1	0	0	1	4	737	0.972296
1	1	1	0	0	4	741	0.977573
1	0	1	0	0	3	744	0.98153
1	1	0	0	1	3	747	0.985488
1	1	1	1	1	3	750	0.989446
0	0	1	1	1	2	752	0.992084
1	1	1	1	1	2	754	0.994723
0	0	1	0	0	1	755	0.996042
0	0	1	0	1	1	756	0.997361
0	1	0	0	1	1	757	0.998681
1	0	1	1	1	0	758	1
0	0	1	1	1	0	758	1
0	1	1	0	0	0	758	1
0	1	1	0	1	0	758	1
0	1	1	1	1	0	758	1
0	1	1	1	1	0	758	1
0	1	1	1	1	0	758	1
1	0	1	0	1	0	758	1
1	1	0	0	1	0	758	1
1	1	1	0	1	0	758	1



## Why limited diversity matters:

Strong Unions (U)	Strong Left Parties (L)	Generous Welfare State (G)	N of Cases
Yes	Yes	Yes	6
Yes	No	No	8
No	No	No	5
No	Yes	????	0 (they don't exist)

Is it strong left parties (L) that cause generous welfare states (G) or is it the combination of strong unions and strong left parties (L•U) that causes generous welfare states (G)? (“•” indicates set intersection—combined causes.)

From a correlational viewpoint, having a strong left party (L) is perfectly correlated with having a generous welfare state (G). A **parsimonious** explanation has been achieved:  $L \rightarrow G$

From a case-oriented perspective, however, all instances of generous welfare state share **two** causally relevant conditions (strong left parties and strong unions) and none of the negative cases have this combination. This pattern suggests a more **complex** explanation:  $L \bullet U \rightarrow G$ .

## Easy versus difficult counterfactual cases:

In the example just provided, the fourth data row is a potential counterfactual case. It is a combination of causal conditions that lacks empirical instances. Is it an “easy” or “difficult” counterfactual case?

Empirically, we know that all instances of generous welfare states exhibit both strong unions and strong left parties (from row 1).

Our background knowledge tells us that the presence of these two conditions should be linked to formation of generous welfare states. That is, we don't expect either weak unions or weak left parties to be linked to the formation of generous welfare states.

We conclude that the combination of weak unions and strong left parties is a **difficult** counterfactual. Therefore, we are **barred** from simplifying  $L \cdot U \rightarrow G$  to  $L \rightarrow G$ , via counterfactual analysis ( $L \cdot U + L \cdot u = L$ ).

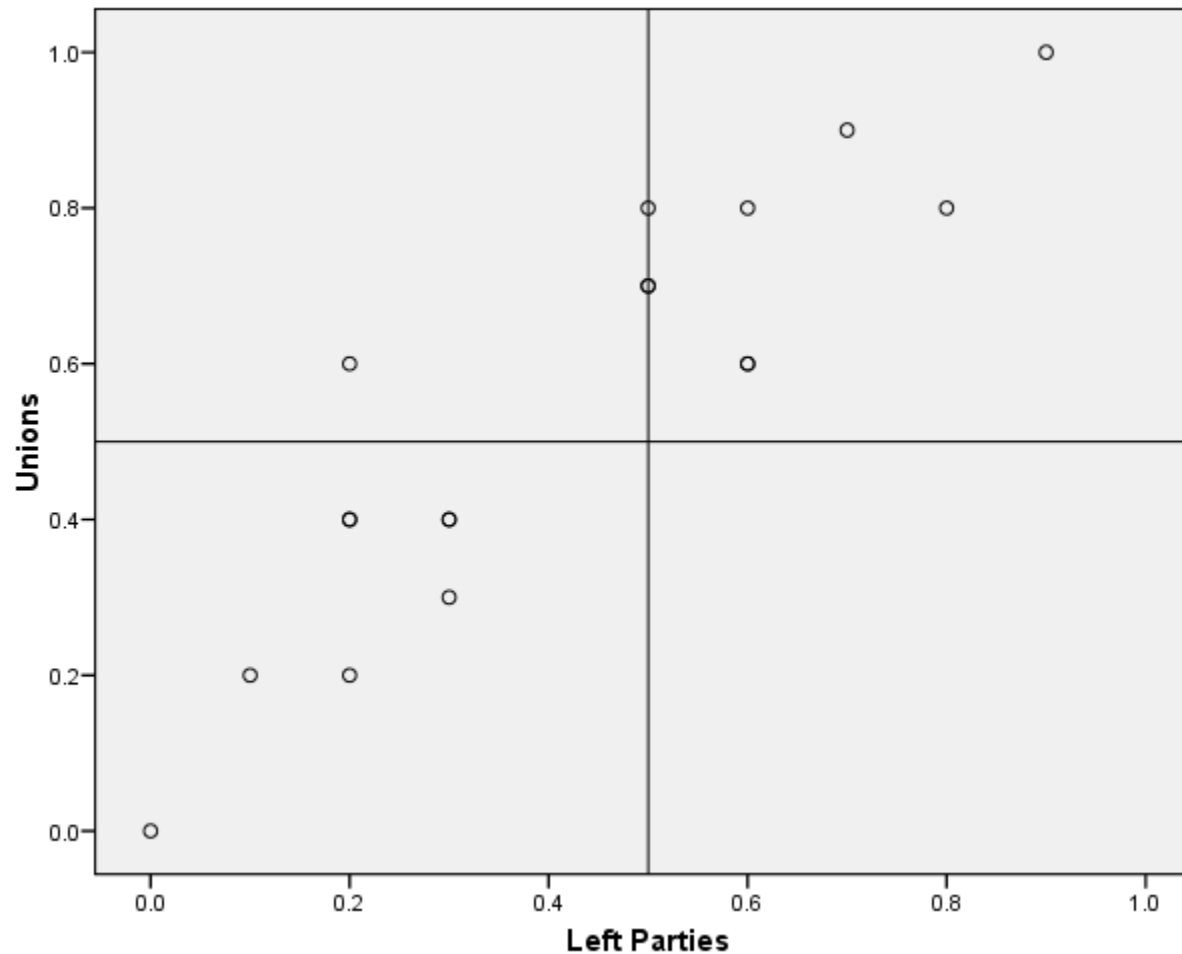
In essence, we have assumed that cases in the fourth row (weak unions combined with strong left party), if they existed, probably would not exhibit generous welfare states. Conventional quantitative analysis makes the opposite assumption, embracing the difficult counterfactual.

Fuzzy set membership scores consistent with previous slide:

<b>Unions</b>	<b>Left</b>	<b>Generous</b>	<b>Country</b>
.0	.0	.0	USA
.2	.1	.1	CANADA
.4	.3	.1	SPAIN
.4	.2	.2	PORTUGAL
.3	.3	.3	FRANCE
.6	.2	.2	ITALY
.6	.6	.5	GERMANY
.8	.6	.5	BELGIUM
.2	.2	.1	SWITZERLAND
.8	.5	.4	AUSTRIA
.7	.5	.5	FINLAND
1	.9	.8	SWEDEN
.8	.8	.8	NORWAY
.9	.7	.7	DENMARK
.6	.6	.4	UK
.4	.2	.2	AUSTRALIA
.4	.3	.2	NEW ZEALAND
.7	.5	.5	IRELAND

Unions  $\geq$  Left  $\geq$  Generous

## Plot of strong unions against strong left parties:



Quadrant #4 is void of cases. There are no cases with strong left parties and weak unions. This pattern parallels what is shown in the crisp 2X2 table.



## Regression analysis of fuzzy-set data:

Dependent Variable: GENEROUS

	<i>Unstandardized Coefficients</i>		<i>Standardized Coefficients</i>	<i>t</i>	<i>Sig.</i>
	B	Std. Error	Beta		
(Constant)	-4.852E-02	.038		-1.267	.225
UNIONS	.152	.145	.169	1.052	.309
LEFT	.784	.156	.808	5.016	.000

a

In short, the regression results indicate that all that really matters is having a strong left party, even though the set-theoretic analysis shows that having strong unions is a necessary-but-not-sufficient condition. (Supersets of the outcome can be viewed as shared antecedent conditions—if they make sense as antecedents.)

The scatterplots for these data are not unusual from the viewpoint of conventional quantitative analysis, other than the fact that they indicate very strong relationships. In other words, there is nothing special about the scatterplots that would make you think that a conventional “net effects” analysis should not be performed.

# Limited diversity in a truth table with four causal conditions:

IMF Pressure	Gov't. Corruption	Strong Unions	Trade Dependence	Austerity Protest
no	no	no	no	no
no	no	no	yes	?
no	no	yes	no	?
no	no	yes	yes	?
no	yes	no	no	no
no	yes	no	yes	no
no	yes	yes	no	?
no	yes	yes	yes	no
yes	no	no	no	?
yes	no	no	yes	?
yes	no	yes	no	?
yes	no	yes	yes	?
yes	yes	no	no	yes
yes	yes	no	yes	yes
yes	yes	yes	no	?
yes	yes	yes	yes	?

## Parsimonious versus complex solutions:

IMF pressure•gov't corruption•weak unions

**complex solution**

IMF pressure

**parsimonious solution**

IMF pressure•gov't corruption

IMF pressure•weak unions

**possible intermediate solutions**

At the left end of the continuum is the complex solution; the right end shows the parsimonious solution. The complex solution is a subset of the parsimonious solution. There are other subsets of the parsimonious solution that are equally valid representations of the evidence in the truth table—"intermediate solutions."

Assume that theoretical and substantive knowledge indicates that it is the presence of these four conditions and not their absence that should be linked to the outcome, IMF protest. This knowledge defines "IMF pressure•gov't corruption•strong unions" as an **easy** counterfactual, yielding solution "IMF pressure•gov't corruption." This same knowledge defines "IMF pressure•no gov't corruption•weak unions" as a **difficult** counterfactual. (This second counterfactual is what is required to produce "IMF pressure•weak unions" as the intermediate solution.)

## Causal conditions linked to successful shaming:

<b>Advice (A)</b>	<b>Commitment (C)</b>	<b>Shadow (S)</b>	<b>Inconvenient (I)</b>	<b>Reverberation (R)</b>	<b>Success (Y)</b>
1	0	1	1	1	1
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	1	1
1	1	1	1	0	0
1	1	1	0	0	1
1	0	0	0	0	1

**Advice (A):** Whether the shamers can substantiate their criticism with reference to explicit recommendations of the regime's scientific advisory body.

**Commitment (C):** Whether the target behavior explicitly violates a conservation measure adopted by the regime's decision-making body.

**Shadow of the future (S):** Perceived need of the target of shaming to strike new deals under the regime--such beneficial deals are likely to be jeopardized if criticism is ignored.

**Inconvenience (I):** The inconvenience (to the target of shaming) of the behavioral change that the shamers are trying to prompt.

**Reverberation (R):** The domestic political costs to the target of shaming for not complying (i.e., for being scandalized as a culprit).

## How this evidence is typical of small-N research:

- The number of cases (10) is more than a handful, but still small enough to permit familiarity with each case. (Two of the eight listed combinations have a frequency of 2.)
- From the viewpoint of conventional quantitative social science, however, the number of cases (10) is very small relative to the number of causal conditions (5). This ratio essentially eliminates the possibility of any form of multivariate statistical analysis.
- If the cases are viewed configurationally, then the analytic prospects seem even more dismal, for there are  $2^5$  logically possible combinations of five causal conditions. There is empirical evidence on only eight of the 32 combinations.
- This pattern of **limited diversity** is characteristic of comparative research and, more generally, of research on naturally occurring social and political phenomena.
- However, causal combinations without cases are potential counterfactual cases. Counterfactual analysis provides an analytic opening.

## Parsimony versus complexity in Stokke's evidence:

A·c·s·i·r +	
A·C·S·i·r +	i +
<u>A·S·I·R</u>	<u>S·R</u>
complex	parsimonious

A·c·s·i·r +		
A·C·S·i·r +	A·i +	i +
<u>A·S·I·R</u>	<u>A·S·R</u>	<u>S·R</u>
intermediate		

In the complex solution, none of the combinations without cases (remainders) is used as a counterfactual case. In the parsimonious solution, any combination without cases that yields a logically simpler solution is incorporated into the solution (i.e., both “easy” and “difficult” counterfactuals have been incorporated). The assumptions are A, C, S, i, R. These assumptions yield the intermediate solution.

The parsimonious solution achieves “perfect” prediction:

success (y)	solution (i + SR)		
	0.0	1.0	
0	4 100.0	0 0.0	4
1	0 0.0	4 100.0	4
Total N	4 8	4	

The solution variable is membership in:  $i + SR$ . (1 = yes; 0 = no.)

According to this formula, shaming works with the behavioral change is not inconvenient for the target to change its behavior or when the target faces the combination of domestic reverberations and the shadow of the future (i.e., the need to strike future deals).

As in the previous examples, it is possible to produce a parsimonious solution that “fits” the data perfectly. The issue is one of soundness. What has been left out?

## Combination A·S·I·R:

1. Causal conditions S and R cannot be removed because they appear in the corresponding parsimonious term at the other end of the continuum.
2. The support of the regime's the scientific advisory body (A) is certainly linked to the success of shaming. This causal condition should be retained.
3. The fact that it is inconvenient for the targets of shaming to change their behavior (I) does *not* promote successful shaming. Thus, inconvenience (I) can be dropped from the combination A·S·I·R because inconvenience of behavioral change to the target of shaming is not central to the success of A·S·R in generating conformity.

$$A \cdot S \cdot I \cdot R + A \cdot S \cdot i \cdot R = A \cdot S \cdot R(I + i) = A \cdot S \cdot R(1) = A \cdot S \cdot R$$

The intermediate combination is **A·S·R**.



## Combination A·C·S·i·r:

1. Condition i (the behavioral change is not inconvenient) cannot be dropped because it appears in the corresponding parsimonious term.
2. Condition A (support from the regime's scientific advisory board) should remain because this condition is clearly linked to the success of shaming.
3. Condition C (the offending behavior clearly violates a prior commitment) also should not be dropped, for this too is something that should only contribute to the success of shaming.
4. Condition S (the violator will need to strike future deals with the regime) is also a factor that should only promote successful shaming.
5. Condition r (absence of domestic reverberations for being shamed) can be removed. Clearly, the presence of domestic reverberation (R) would promote successful shaming.

The intermediate combination is **A·C·S·i**.

## Combination A·c·s·i·r:

1. Condition i must be retained because it appears in the corresponding parsimonious term.
2. Condition A is retained as well, for the reasons stated previously.
3. Condition r (absence of domestic reverberations) can be removed, as it was from the previous combination, for the same reason provided.
4. Condition c (absence of violation of a commitment) can be removed, for surely these instances of successful shaming would still have been successful if there had been an explicit violation of a commitment (C).
5. Condition s (absence of a need to strike future deals with the regime) can be safely removed because only its presence (S) should contribute to the success of shaming.

The intermediate term is **A·i**.

## Boolean simplification:

These three intermediate terms can be joined into a single equation:

$$A \cdot S \cdot R + A \cdot C \cdot S \cdot i + A \cdot i \longrightarrow Y$$

which can then be simplified to:

$$A \cdot S \cdot R + A \cdot i \longrightarrow Y$$

because the term  $A \cdot C \cdot S \cdot i$  is a subset of the term  $A \cdot i$  and is thus logically redundant. (All cases of  $A \cdot C \cdot S \cdot i$  are also cases of  $A \cdot i$ .) These results indicate that there are two paths to successful shaming: (1) support from the regime's scientific advisory body (A) combined with the need to strike future deals (S) and domestic reverberations for being shamed (R), and (2) support from the regime's scientific advisory body (A) combined with the fact that the behavioral change is not inconvenient (i).

## Counterfactual analysis using QCA:

Should contribute to Y when cause is:

Present

Absent

Present or Absent

Causal Condition:

A	0	0	0
C	0	0	0
S	0	0	0
I	0	0	0
R	0	0	0

This QCA dialogue box, in effect, makes it possible for the user to input theoretical and substantive knowledge, with respect to the links between causal conditions and the outcome. The impact is to permit the use of “easy” counterfactual cases, which in turn make it possible to remove counterintuitive elements from the complex solutions (provided that these removals do not violate the parsimonious solution).

# fsQCA results: Stokke data

## Complex Solution:

	raw coverage	uni que coverage	consi stency
	-----	-----	-----
A*S*I*R	0. 500000	0. 500000	1. 000000
A*c*s*i *r	0. 250000	0. 250000	1. 000000
A*C*S*i *r	0. 250000	0. 250000	1. 000000

sol uti on coverage: 1. 000000

sol uti on consi stency: 1. 000000

## Parsimonious Solution:

	raw coverage	uni que coverage	consi stency
	-----	-----	-----
i	0. 500000	0. 500000	1. 000000
S*R	0. 500000	0. 500000	1. 000000

sol uti on coverage: 1. 000000

sol uti on consi stency: 1. 000000

## Intermediate Solution:

	raw coverage	uni que coverage	consi stency
	-----	-----	-----
A*i	0. 500000	0. 500000	1. 000000
A*S*R	0. 500000	0. 500000	1. 000000

sol uti on coverage: 1. 000000

sol uti on consi stency: 1. 000000

## Assumptions of intermediate solution:

A	C	S	I	R	Y
0	0	0	1	0	0
1	0	0	0	0	1
1	0	0	1	0	0
1	0	0	1	1	0
1	0	1	1	1	1
1	1	1	0	0	1
1	1	1	1	0	0
1	1	1	1	1	1
1	0	0	0	1	*
1	0	1	0	0	*
1	0	1	0	1	*
1	1	0	0	0	*
1	1	0	0	1	*
1	1	1	0	1	*

The *intermediate* solution assumes that if instances of the rows with “\*” in the outcome column (y) could be found, they would display the outcome (successful shaming).

# Assumptions of parsimonious solution:

a	c	s	i	r	y
0	0	0	1	0	0
1	0	0	0	0	1
1	0	0	1	0	0
1	0	0	1	1	0
1	0	1	1	1	1
1	1	1	0	0	1
1	1	1	1	0	0
1	1	1	1	1	1
0	0	0	0	0	*
0	0	0	0	1	*
0	0	1	0	0	*
0	0	1	0	1	*
0	0	1	1	1	*
0	1	0	0	0	*
0	1	0	0	1	*
0	1	1	0	0	*
0	1	1	0	1	*
0	1	1	1	1	*
1	0	0	0	1	*
1	0	1	0	0	*
1	0	1	0	1	*
1	1	0	0	0	*
1	1	0	0	1	*
1	1	1	0	1	*

The *parsimonious* solution assumes that if instances of the rows with “\*” in the outcome column (y) could be found, they would display the outcome (successful shaming).

## Another way to think about counterfactual analysis:

When qualitative researchers study an outcome, they focus primarily on the positive cases. This focus derives from an interest in “how” things happen. Negative cases, because they lack the outcome, are not very useful when it comes to explaining how an outcome comes about. In statistical parlance, qualitative researchers commit the crime of “selecting on the dependent variable” when they focus on positive cases.

Using the Stokke data, the positive cases are:

<b>Advice (A)</b>	<b>Commitment (C)</b>	<b>Shadow (S)</b>	<b>Inconvenient (I)</b>	<b>Reverberation (R)</b>	<b>Success (Y)</b>
1	0	1	1	1	1
1	1	1	1	1	1
1	1	1	0	0	1
1	0	0	0	0	1



When qualitative researchers examine positive cases, they search for causal conditions that “make sense” as conditions that “should be” linked to the outcome. They are likely to pass over those that don’t make sense. In other words, they use their background knowledge to identify the relevant causal conditions in each case. If a contributing cause is absent, it is considered irrelevant (as opposed to being relevant in its absence). This focus on “contributing cause is present versus irrelevant” transforms the data as follows:

<b>Advice (A)</b>	<b>Commitment (C)</b>	<b>Shadow (S)</b>	<b>Inconvenient (I)</b>	<b>Reverberation (R)</b>	<b>Success (Y)</b>
1	-	1	-	1	1
1	1	1	-	1	1
1	1	1	0	-	1
1	-	-	0	-	1

The dashes ( - ) indicate that a condition has been designated irrelevant because it is inconsistent with the researchers substantive and theoretical knowledge. In other words, the condition does not make sense as a contributing cause. For example, a behavioral change that is “inconvenient” is unlikely to be a contributing cause when it comes to successful shaming.

This table can be represented as an equation and then simplified using Boolean algebra:

$$A \cdot S \cdot R + A \cdot C \cdot S \cdot R + A \cdot C \cdot S \cdot i + A \cdot i = \text{successful shaming}$$

“.” indicates set intersection (combined conditions)

“+” indicates set union (alternate combinations)

$A \cdot C \cdot S \cdot R$  is a subset of  $A \cdot S \cdot R$  and is therefore redundant.

$A \cdot C \cdot S \cdot i$  is a subset of  $A \cdot i$  and also is redundant.

The simplified equation is:

$$A \cdot S \cdot R + A \cdot i = \text{successful shaming}$$

Thus, this formalization of conventional case-oriented practice produces results that are the same as QCA's intermediate solution.

## The negative cases:

These same principles of qualitative analysis can be applied to Stokke's instances of unsuccessful shaming. First, consider the table showing the four combinations linked to unsuccessful shaming:

<b>Advice (A)</b>	<b>Commitment (C)</b>	<b>Shadow (S)</b>	<b>Inconvenient (I)</b>	<b>Reverberation (R)</b>	<b>Success (Y)</b>
1	0	0	1	0	0
1	0	0	1	1	0
0	0	0	1	0	0
1	1	1	1	0	0

Next, observe the transformation of “presence versus absence” to “contributing versus irrelevant” (notice that these codings are the opposite of the codings for the positive cases):

<b>Advice (A)</b>	<b>Commitment (C)</b>	<b>Shadow (S)</b>	<b>Inconvenient (I)</b>	<b>Reverberation (R)</b>	<b>Success (Y)</b>
-	0	0	1	0	0
-	0	0	1	-	0
0	0	0	1	0	0
-	-	-	1	0	0

Conversion of the second table to equation form yields:

$$c \cdot s \cdot l \cdot r + c \cdot s \cdot l + a \cdot c \cdot s \cdot l \cdot r + l \cdot r = \text{shaming failed}$$

“.” indicates set intersection (combined conditions)

“+” indicates set union (alternate combinations)

$c \cdot s \cdot l \cdot r$  is a subset of both  $c \cdot s \cdot l$  and of  $l \cdot r$  and is therefore redundant.

$a \cdot c \cdot s \cdot l \cdot r$  is a subset of both  $c \cdot s \cdot l$  and of  $l \cdot r$  and also is redundant.

The simplified equation is:

$$c \cdot s \cdot l + l \cdot r = \text{shaming failed}$$

Set-theoretic analysis is inherently asymmetrical. The causal conditions linked to the presence of an outcome (shaming succeeded) are rarely the exact reverse of those linked to its opposite (shaming failed).

# Limited diversity in a typical large-N social science data set:

College	Parental Income	AFQT Score	Married	Children	Freq.	Cum Freq	Cum %
0	0	0	0	0	327	327	0.431398
0	0	0	0	1	154	481	0.634565
0	0	0	0	0	65	546	0.720317
1	0	0	0	0	57	603	0.795515
0	0	0	1	0	41	644	0.849604
1	0	0	1	1	24	668	0.881266
0	1	0	0	0	14	682	0.899736
1	0	0	1	0	13	695	0.916887
0	1	0	1	1	10	705	0.930079
1	1	0	0	0	10	715	0.943272
1	0	0	0	0	7	722	0.952507
1	0	1	1	1	6	728	0.960422
0	1	0	1	1	5	733	0.967018
1	1	0	1	1	4	737	0.972296
1	1	1	0	0	4	741	0.977573
1	0	1	0	0	3	744	0.98153
1	1	0	1	1	3	747	0.985488
1	1	1	1	1	0	750	0.989446
0	0	1	1	1	1	752	0.992084
1	1	1	1	1	1	754	0.994723
0	0	1	0	0	1	755	0.996042
0	0	1	0	1	1	756	0.997361
0	1	0	0	1	1	757	0.998681
1	0	1	1	1	0	758	1
0	0	1	1	1	0	758	1
0	1	1	0	0	0	758	1
0	1	1	0	1	0	758	1
0	1	1	1	1	0	758	1
0	1	1	1	1	0	758	1
1	0	1	0	0	1	758	1
1	1	0	0	0	1	758	1
1	1	1	1	0	1	758	1

Logistic regression of poverty avoidance on AFQT scores, parental income, years of education, martial status, and children:

	<b>B</b>	<b>S.E.</b>	<b>Sig.</b>	<b>Exp(B)</b>
<b>AFQT (z score)</b>	.391	.154	.011	1.479
<b>Parental Income (z score)</b>	.357	.154	.020	1.429
<b>Education (z score)</b>	.635	.139	.000	1.887
<b>Married (yes = 1, 0 = no)</b>	1.658	.346	.000	5.251
<b>Children (yes = 1, 0 = no)</b>	-.524	.282	.063	.592
<b>Constant</b>	1.970	.880	.025	7.173

Chi-Squared = 104.729, df = 5

It is possible to use this equation to predict the log odds of poverty avoidance for cases in all thirty-two sectors of the vector space defined by the five causal conditions. However, as shown already, most of these sectors are void or virtually void of cases.

What makes these predictions possible? Simplifying assumptions are built into the estimation technique (e.g., assumptions of linearity and additivity).

## Conclusions:

1. Limited diversity is a characteristic feature of naturally occurring social phenomena.
2. The resolution of the problem of limited diversity involves the use of counterfactual analysis in some way.
3. In case-oriented comparative research, the resolution of the problem of limited diversity is knowledge and theory dependent. “How” this happens in case-oriented research (Achen’s query) is through the incorporation of “easy” counterfactuals.
4. In order to define “easy” counterfactuals, researchers must apply their substantive and theoretical knowledge to the “remainder” combinations. In practice, this allows them to craft an intermediate solution, situated between the “most complex” and “most parsimonious” QCA solutions. It is necessary to maintain the subset relationship among possible solutions along the complexity/parsimony continuum.
5. In quantitative research, the problem of limited diversity is also addressed through assumptions. However, these assumptions (e.g., linearity and additivity) are usually invisible to most users.
6. The incorporation of background knowledge through “easy” counterfactuals is central to case-oriented comparative research. This process is made explicit in QCA.