Working Paper 2016-85

The set theoretic analysis of probabilistic regularities with fsQCA, QCApro and CNA: Exploring possible implications for current "best practice"

Barry Cooper
Durham University

Judith Glaesser
Durham University

Abstract: Lieberson (1997a,b) and Goldthorpe (2007, 2016), both eminent critics of QCA, argue that social life is characterised in part by probabilistic processes. Both claim that QCA is ill-equipped to respond to this aspect of social reality (Goldthorpe, 2007, 2016, Lieberson, 1991, 2001). In fact, Ragin, in his earlier work (e.g. 1987), does employ probabilistic language to describe differences in outcome by configuration and, later (2000), explicitly discusses the role of randomness in the processes generating outcomes. Notwithstanding this background, there has been little attention paid to the consequences of such *generative* randomness – as opposed to measurement and sampling error – for the practices of QCA and related set theoretic techniques such as CNA. In this exploratory paper we discuss some of these consequences.
The paper has these elements:

1. We discuss what Lieberson, on the one hand, and Ragin, on the other, have had to say about probability, chance and randomness in social life and analysis. We use this discussion to develop and set out the approach we use to simulate generative randomness in this paper.
2. We then discuss and analyse invented datasets generated by a mix of deterministic and random processes (the latter simulated by the use of a lottery to allocate scarce university places) to bring out the problems that generative randomness might raise for set theoretic analysis.
   i. We begin with datasets that standard QCA practice can handle adequately. We use our discussion of these, first, to note the limitations of purely regularities-based analyses and, second, to suggest a way of reporting the results of QCA in situations where it is known or suspected that the generative mechanism includes a random process that leads to "truly" probabilistic outcomes. We also argue here that there is an important ambiguity in the way the term "probability" has been used by both Lieberson and Ragin, of which any reporting of QCA solutions in terms of probabilistic outcomes must take account.
   ii. We then consider a further dataset, also partly reflecting generative randomness, that, as a consequence of its solution combining very low suf-consistency with perfect suf-coverage, causes a number of difficulties for typical practice. In turn, we discuss how fsQCA (Ragin and Davey, 2014), QCApro for R (Thiem, 2016a,b; developed and expanded from QCA for R, Thiem and Duşa, 2013a,b) and Baumgartner's CNA (Ambuehl et al., 2014, Baumgartner, 2009, Baumgartner and Thiem, 2015a) behave in

this context. Discussing CNA, we give some space to the problems our partly non-deterministic set-up causes for its capacity to analyse causal chains (Baumgartner and Epple, 2014).

3. Finally, we summarise the lessons of our exploratory analyses, noting some outstanding questions.

Keywords: chance, generative randomness, probabilistic processes, lottery.
Methods: csQCA, CNA

COMPASSS Working Paper 2016-85
Available at http://www.compasss.org/wpseries/CooperGlaesser2016.pdf
COMPASSS Working Paper Series Managing Editor, Claude Rubinson

The set theoretic analysis of probabilistic regularities with fsQCA, QCApro and CNA: Exploring possible implications for current "best practice"

Barry Cooper & Judith Glaesser

Durham University, UK

Abstract

Lieberson (1997a,b) and Goldthorpe (2007, 2016), both eminent critics of QCA, argue that social life is characterised in part by probabilistic processes. Both claim that QCA is ill-equipped to respond to this aspect of social reality (Goldthorpe, 2007, 2016, Lieberson, 1991, 2001). In fact, Ragin, in his earlier work (e.g. 1987), does employ probabilistic language to describe differences in outcome by configuration and, later (2000), explicitly discusses the role of randomness in the processes generating outcomes. Notwithstanding this background, there has been little attention paid to the consequences of such *generative* randomness – as opposed to measurement and sampling error – for the practices of QCA and related set theoretic techniques such as CNA. In this exploratory paper we discuss some of these consequences.

The paper has these elements:

1. We discuss what Lieberson, on the one hand, and Ragin, on the other, have had to say about probability, chance and randomness in social life and analysis. We use this discussion to develop and set out the approach we use to simulate generative randomness in this paper.
2. We then discuss and analyse invented datasets generated by a mix of deterministic and random processes (the latter simulated by the use of a lottery to allocated scarce university places) to bring out the problems that generative randomness might raise for set theoretic analysis.
   i. We begin with datasets that standard QCA practice can handle adequately. We use our discussion of these, first, to note the limitations of purely regularities-based analyses and, second, to suggest a way of reporting the results of QCA in situations where it is known or suspected that the generative mechanism includes a random process that leads to "truly" probabilistic outcomes. We also argue here that there is an important ambiguity in the way the term "probability" has been used by both Lieberson and Ragin, of which any

reporting of QCA solutions in terms of probabilistic outcomes must take account.

ii. We then consider a further dataset, also partly reflecting generative randomness, that, as a consequence of its solution combining very low suf-consistency with perfect suf-coverage, causes a number of difficulties for typical practice. In turn, we discuss how fsQCA (Ragin and Davey, 2014), QCApro for R (Thiem, 2016a,b; developed and expanded from QCA for R, Thiem and Duşa, 2013a,b) and Baumgartner's CNA (Ambuehl et al., 2014, Baumgartner, 2009, Baumgartner and Thiem, 2015a) behave in this context. Discussing CNA, we give some space to the problems our partly non-deterministic set-up causes for its capacity to analyse causal chains (Baumgartner and Epple, 2014).

3. Finally, we summarise the lessons of our exploratory analyses, noting some outstanding questions.

The paper is long and detailed. We therefore offer some suggested routes through it. The shortest and simplest version, focusing just on fsQCA, would comprise sections 1-5 plus 9. A slightly longer version, comprising sections 1-6 and 9, adds a discussion focused on the R-based QCApro. For the reader interested in CNA, and the analysis of causal chains, sections 7 and 8 should also be read.

1: Introduction

The debate on the capacity of set theoretic methods to deliver causal knowledge continues to develop in breadth and depth. A somewhat neglected issue, however, is how the various instantiations of set theoretic methods in software and/or as best practice rules behave when confronted by datasets whose generation reflects not just sampling and measurement error but rather, *as part of the underlying causal structure generating the outcome*, some random process. There has been some discussion of this issue (see, e.g., Ragin, 1995; 2000:108-9, 222-3, 226-9). However, many scholars aim to establish deterministic (i.e., non-probabilistic) accounts of the causal paths to some outcome, with any failure to do so being understood by reference to omitted causal conditions or measurement and/or sampling error. The indices of consistency and coverage employed to assess the degree of fit to perfect relations of sufficiency and necessity, for example, are often seen as addressing the problem of "noise" in data (e.g. Baumgartner and Epple, 2014). Instead, we want to explore these indices in situations where there is *generative* randomness involved in the production of some outcome.

We will develop our discussion by considering invented datasets generated partly by such random processes but, to keep the treatment manageable, *not* affected in any way by sampling or measurement error. By random generative processes, we refer to processes similar to those than can be simulated by the operation of roulette wheels, dice throws or lotteries[1].

Our interest in this topic has been reinforced by the fact that two eminent critics of QCA have written extensively on the role of probabilistic processes in social life (Goldthorpe, 2007, 2016, Lieberson, 1997a,b) and, furthermore, have claimed that the advocates of logical approaches such as QCA have paid insufficient attention to the problems such processes produce for Boolean approaches (Goldthorpe, 2007, 2016, Lieberson, 1991, 2001). Lieberson (2001), reviewing Ragin (2000), claims that the author

> seems to be forced to resort to using probabilistic procedures to determine whether there is a necessary condition or a sufficient condition. This not only sounds like an oxymoron, it is. We have conditions deemed to be "almost always necessary" (p. 272) and "almost always sufficient" (p. 279). He feels free to resort to a probabilistic view because this allows for measurement and translation imprecisions. He is unwilling to entertain the possibility that the deviations are not errors, but rather that the conditions are not truly necessary or not truly sufficient, but are simply probabilistic statements. He provides no procedure for distinguishing between the two centrally different reasons. What is crucial here is that when you begin to think of a concept such as a necessary condition being one which is not fully so, then you are shifting over to the basic way many conventional social scientists think: conditions that increase or decrease the likelihood of some outcome.

Lieberson (2004) repeated this view that QCA does not adequately allow for "chance and probabilistic processes". Goldthorpe (2016: 53), arguing that social analysis should primarily focus on establishing and explaining "probabilistic regularities", also argues that usages such as "almost sufficient" are oxymoronic. He claims (53-54) that QCA can make no allowance for "essential chance" (Monod, 1971), i.e. of the possibility that processes in the social world are not fully deterministic in nature.

We can understand the unhappiness of some social scientists with such unconventional usages but will argue that there is a defensible way to avoid the apparent difficulties of conjoining terms like "usually" and "almost" with "necessary" and "sufficient". Our approach is related to that used in an early paper by Ragin (1995) but we will also take account of the arguments of Lieberson and Goldthorpe concerning "probabilistic processes"[2]. However, given the critical comments on Ragin's work just referred to, we should, before moving on, consider what he has written about probability and randomness in the context of Boolean analysis.

2: Ragin on QCA, Randomness, Probability and Contingency

Most of the discussion of randomness in the QCA literature has focused on measurement (including calibration) and sampling error (e.g. Skaaning, 2011, Thiem, Spöhel and Dușa, 2016)[3] rather than the ways generative random processes might interact with deterministic causation in producing outcomes[4]. There has, however, been some discussion of "probabilistic processes" in the context of outcome generation. It is Ragin (2000), rather than others who have developed his work[5], who has had most to say on this. However, before discussing his remarks there, we need to consider the more conventional references he made to "probabilities" – as "relative frequencies" – in his earlier work. In this work, Ragin followed conventional quantitative researchers in referring to the "probabilities" of various outcomes. Ragin et al. (1984: 229-230), discussing a Boolean analysis of discrimination in employment, note:

> Once individuals are aggregated into similar situations, however, their scores on the dependent variable, promotion, are aggregated into the *probability of promotion* for individuals who share a given combination of values. Thus, it is necessary to assign these *probabilities* to pass or fail conditions or to trichotomize them into pass, fail, or neither pass nor fail. (our italics)

Ragin (1987: 117) talks of "probabilities of success … for each causal combination". Ragin and Bradshaw (1991) use similar terminology. This interpretation of QCA in terms of probabilities also characterises Ragin (1995), where he analyses outcomes by configuration in terms of three degrees of likelihood: uniform (all cases achieve the outcome), likely and possible[6]. In this framework, "probabilistic outcomes are intermediate between uniform and possible outcomes". In these papers, then, Ragin employs "probability" in a conventional manner. Relative frequencies, derived from the study of some sample, are used as estimates of the "probability" of individuals, with certain characteristics, achieving some outcome, *ceteris paribus*[7].

In later work, Ragin (2000: 108) does explicitly note, alongside those of complexity, measurement error and the "misconstructing" of populations on QCAs, the effects of "random factors":

> Consider again the example of strikes that occur in response to the introduction of technology opposed by workers. Suppose a charismatic worker makes a stirring speech on the behalf of the new technology and convinces workers not to go on strike, but instead to give the new production techniques a try. Suppose a flood closes the plant for a month, and then workers flock back to work, eager for overtime bonuses, when the waters recede. Suppose anarchists have infiltrated the union leadership, and the rank-and-file members refuse to follow any of their recommendations, no matter how sensible they may seem. There are many such minor, obscure, or random factors that might interfere with the expected connection between a cause and an effect. It is virtually impossible to construct social scientific models that take account of every possible factor that might influence some action or outcome.

And later, he refers to "randomness" (223):

> The third issue is randomness. Even with precise measures and well-reasoned translations to fuzzy membership scores, the researcher is still likely to confront ill-fitting data. The social world is multicausal, and a variety of causal conditions unique to each case, including exogenous events like snowstorms and flu epidemics, affect every outcome. Thus, cases almost never plot exactly where they "should." Consider again the hypothetical analysis of communities with drug problems. Some communities may be more resistant or more vulnerable to drug problems than others, depending on their histories. Community-specific differences may be 'very difficult to detect, much less to operationalize and measure in a meaningful cross-case manner. Thus, factors relevant to each specific community are likely to be left out of the analysis altogether, increasing the lack of fit between membership in the outcome and membership in relevant causal conditions.

Here "randomness" seems to be understood more as coincidence or contingency[8], concerning factors external to the explored model, than as the sort of randomness characterising such activities as dice-throwing. "Randomness" here might be seen as an example of a Cournot Effect, where two or more apparently independent causal chains converge to produce a chance outcome (Boudon, 1986: 175)[9]. Boudon gives the oft-used example of a slate falling from a roof and hitting a passer-by. While the slate falling and the man passing are open to causal explanation, the fact that the two events converge is not causally determined[10]. Since this account of "randomness" does provide one basis for arguing, with Lieberson and Goldthorpe[11], that there are random generative processes in the social world or at least, but more weakly, that there are processes that we will effectively have to treat as if they are random, given limited theoretical knowledge and limited access to data, it is worth noting Boudon's formalisation of a Cournot Effect. He writes (178-9):

> Chance is therefore not nothing. It is a particular form that sets of cause/effect linkings as perceived by a real observer can take on. Some of them have a total form of order …. Others have a partial form of order …. Others contain contingent links (the series "A causes B, which causes C" occurs at the same time as "P causes Q, which causes R") but it is impossible to decide whether the synchronization is really between B and P, B and Q or C and Q. It is therefore impossible to tell whether event BP, BQ or CQ will necessarily be brought about. And the three events can have very different consequences. So there is such a thing as chance …. We must see chance not as a substance, a variable or a set of variables, but as a structure which is characteristic of certain sets of causal chains as perceived by an observer.

We have here one way of justifying the claim that we need to take chance seriously. Although the example concerns a singular outcome, it can be seen that, in the context of a concern with the quasi-regularities that characterise such matters as the relations between socio-economic factors and educational and/or health outcomes, we could, by applying a similar argument to each individual case, provide an argument that, as Goldthorpe (2016) claims, social scientists will typically be faced with "probabilistic regularities" as their explanandum. Boudon, employing Cournot's insight, provides a viable account of Ragin's (2000) "randomness".

It might be objected that the processes being described are not in themselves "probabilistic" in the same way that quantum processes are thought to be in physics[12]. Boudon's phrase "as perceived by an observer" hints at this problem. There will, in fact, be degrees of (un)predictability here. One scholar might argue that the causes of health problems (one causal series) are "independent of but interfere with" the causal processes operating in the educational sphere (a second causal series): a child expected to achieve well in fact does not due to illness. However, another might reasonably respond that social class plays a role in both of these causal series and, to this extent, the assumption of independent series is one to be questioned. Claims of independence will be contestable. To avoid such contestability, we employ in this paper a more clear-cut approach to generative randomness. We assume that there are devices that can be taken to simulate random processes. We use the device of a lottery but die-throwing would serve equally well[13]. In our simulations, we embed our device in an otherwise deterministic closed system with *known to us* causal conditions, processes and outcomes. We explore the behaviour of set theoretic methods in such settings.

Our purpose, then, is to discuss how set theoretic approaches cope with invented datasets whose generation incorporates a known random process, with a lottery representing that process. Our discussion has practical relevance. Since there are lotteries in the world then, assuming that a lottery can be taken to represent randomness, there are random elements in at least some areas of social life, i.e. those that include lotteries as part of their operation. These include education as part of admissions procedures and, of course, gambling[14]. However, our motivation for developing these datasets is our wish to explore what we can learn from analysing them that might be of more general use to those interested in currently recommended procedures for sufficiency and necessity-based causal analyses. The lottery is being used here as a convenient device to simulate randomness.

Although we have given the paper a realistic flavour by discussing an imagined educational context, we think that nothing in our discussion hinges on this. We develop our arguments, as we have done elsewhere (Cooper and Glaesser, 2011, 2015a,b, 2016), by undertaking a range of analyses of *invented* datasets. A key point is that we, as the inventors, will know the causal relations and "random" processes that generated the data. However, we initially simulate the work of a scholar who doesn't share our designer's knowledge and must therefore focus on observable regularities. We thereby concentrate on the most basic form of set theoretic analysis, that of regularities. We also restrict ourselves to parsimonious solutions of truth tables. This keeps our discussion simpler. It is also the case that some scholars (e.g. Baumgartner (2015) and Baumgartner & Thiem (2015b) argue that, given a causal focus, only parsimonious solutions can be relied upon to produce valid results. Ragin (2008), however, suggests that parsimonious solutions should be used with caution. As it happens, we will be discussing models of invented population-level datasets where we *assume* we have access to data on all possible cases, and we think that our arguments here are independent of those concerning the choice between parsimonious, intermediate and conservative/complex solutions (see Cooper and Glaesser, 2015b, Schneider and Wagemann, 2012, 2015, and Thiem 2015a for recent discussions of the implications of limited diversity).

The first datasets discussed are designed so that conventional thresholds for consistency with sufficiency are easily met. The discussion of these enables us to present an appropriate form of set theoretic description for regularities generated partly by random processes. Subsequently we discuss a more difficult but crucial case where conventional thresholds for quasi-sufficiency are not met but indices of coverage are high. We argue that, in this consistency/coverage context, the consequences of generative randomness for standard advice on the use of consistency and coverage measures need to be more widely considered and understood.

3: Three invented worlds with high suf-consistency and perfect suf-coverage for an outcome

We assume initially that a scholar has no knowledge of the processes and mechanisms linking conditions and outcome, i.e. of the generative mechanisms and processes producing *by our design* the analysed dataset. Clearly, this pure separation of knowledge of regularities and generative mechanisms does not characterise actual research practice. Even in choosing factors to analyse, researchers draw on theoretical knowledge of how or why these might affect some outcome. Some writers on Boolean approaches do, however, argue that the causal relevance of factors can, in some circumstances, be established by analysing regularities (e.g., Baumgartner, 2008)[15]. Others act as if this were the case, paying little attention to generative mechanisms[16]. Our reason for focusing initially on regularities is methodological. This enables us to note that identical solutions can arise from analysing regularities generated by different mechanisms. In particular, we are able to show why "random" generative processes complicate recommended standard practices. We mainly employ upper / lower case notation to indicate the presence / absence of binary conditions and outcomes. We abbreviate configurations containing logical AND such as "A AND B AND c" to simply ABc, and use + for logical OR. Occasionally, for clarity, we indicate that a condition has the value 1 or 0 explicitly. Finally, where the output from software uses an asterisk to indicate AND we sometimes retain this for formatting purposes.

We employ initially a simple example to prepare the ground for discussing a more complex model. Consider Figure 1. Some single factor X is thought to be causally relevant for Y, but our scholar has no knowledge of what goes on within the black box.

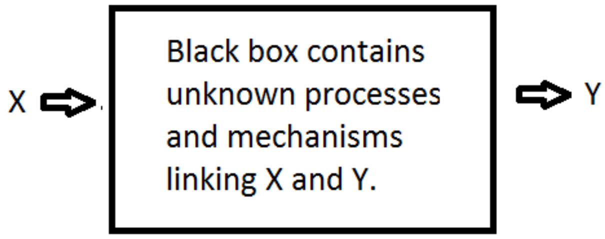

Figure 1: Y=f(X)

The X-Y truth table (with 2000 cases) is found to be Table 1. Using the fsQCA software with a 0.9 consistency threshold, the scholar finds X is quasi-sufficient (suf-consistency=0.9) and necessary for Y (nec-consistency=1.0)

Table 1: truth table for Y=f(X)

X	Number of cases *with* outcome: Y	Number of cases *without* outcome: y	Consistency for Y	Number of cases
0	0	1000	0	1000
1	900	100	0.9	1000

Many would be happy with the indices here (0.9 / 1.0). Others might want to find other causal conditions that, combined with X, might improve suf-consistency. They might add conditions to the model without explicitly addressing the processes within the black box. Others might aim to “explain” the 0.9 level of consistency. To achieve this, they would, we believe, need more knowledge of processes in the black box. The obvious way of gaining this would be to use process tracing (Collier, 2011, George and Bennett, 2005, Glaesser, 2015) to ascertain the nature of the processes connecting X and Y.

We now make the discussion more substantive by assuming X is some measure of high academic ability and Y the achievement of a university degree. We can then construct three expanded versions of the dataset by introducing alternative sets of invented factors and processes intervening between X and Y to create the X-Y regularities. Each of these three datasets reflects an alternative possible set of processes and mechanisms. Importantly, one of these datasets *by our design* includes effects generated partly by a deliberately random process (a lottery). To keep our discussion simpler, we assume that these are population level data, ensuring we have no sampling error, and also that there is no measurement/calibration error. Assume then there are three invented societies under study (each with a relevant population of 2000). While the processes linking X and Y differ, all three have the X-Y regularities in Table 1. We can model these differing processes by adding intervening factors to the truth table. These are shown in Table 2 (society A), Table 3 (society B) and Table 4 (society C). In society A, all cases with X (and only these) achieve S (a secondary school diploma), and 90% of those with S achieve Z (a higher level of pass). All of those with Z (and only these) enter university and achieve a degree. In society B all of those with X=1 (and only these) gain S. All of those with S (and only these) enter the university, but 10% drop out before achieving their degrees. Finally, in society C, all of those with X=1 (and only these) achieve the secondary school pass, S, and these individuals (who all want to attend university) must enter a lottery (used to distribute scarce places to 90% of the lottery participants), the winners of which gain a place in the university. The latter all achieve a degree. Only individuals with S=1 are allowed access to this lottery.

Table 2: Society A

X (high academic ability)	S (Secondary school diploma: pass)	Z (Secondary school diploma: pass at higher level)	Number achieving a university degree: Y	Number *not* achieving a university degree: y	Consistency for Y (achieving a university degree)
0	0	0	0	1000	0
1	1	0	0	100	0
1	1	1	900	0	1

Table 3: Society B

X (high academic ability)	S (Secondary school diploma: pass)	E (enters university: but 10% drop out)	Number achieving a university degree: Y	Number *not* achieving a university degree: y	Consistency for Y (achieving a university degree)
0	0	0	0	1000	0
1	1	1	900	100	0.9

Table 4: Society C

X (high academic ability)	S (Secondary school diploma (pass))	L (Entered for a Lottery with 90% winners)	Number achieving a university degree: Y	Number *not* achieving a university degree: y	Consistency for Y (achieving a university degree)
0	0	0	0	1000	0
1	1	1	900	100	0.9

Crucially, in each of these societies, if the model Y=f(X) is analysed, X will be found to be necessary and (quasi-)sufficient for Y with identical suf-coverage (1.0) and suf-consistency (0.9) figures. A regularities-based analysis of just Y=f(X) does not distinguish between these three scenarios[17]. But how should a scholar respond to the case of Society C, assuming now s/he does, at a second stage, gain knowledge of the mechanism involving a lottery that links

X and Y here? We think that, if the analysis is simply of Y=f(X), the following expressions capture Society C:

Equation 1

{X=1} = > {Probability(any such individual gets a degree)=0.9}.

Equation 2

{Having a degree} => {X=1}

A key point is that the probability in Equation 1 correctly describes the individual case as well as the configuration. Each case with X=1 in society C enters the lottery and has a 0.9 probability of achieving a degree. This probability becomes an objective characteristic of these cases in the same way that the probability of 1/6 of obtaining a six is an objective feature of a tumbled die. It is a feature of the invented world. It is also the case, of course, that if the researcher were randomly to pick a case (late enough in the life cycle) from within the set having X=1 in society C then the probability of this case having a degree would be 0.9 – in the alternative sense that if this sampling were repeated (with replacement) then, in the long run, 90% of picks would have a degree. The "dispositional" probability at the level of the case and the "relative frequencies" probabilities match. This identity of the two arises from the role of the lottery: society C is *really* partly probabilistic. It is *because* the dispositional *probability* of any case with X=S=1 gaining a degree is 0.9 in this world itself that the *proportion* gaining a degree is found to be 0.9 by the researcher investigating Y=f(X)[18].

The situation in societies A and B is different. While a QCA of Y=f(X) produces the same result as society C, it is not so clearly the case that the *dispositional* probabilities at the individual case level match the "relative frequency" probabilities at the configurational level. A researcher randomly picking cases from the set with X=1 will again find that, in the long run, 90% gain degrees. But it may not be the case, in these two societies, that each individual with X=1 has a 0.9 chance in the objective sense that they do in society C of gaining a degree. If we assume, for illustration, that the link in society A between X, S and Z involves no randomness but is, in fact, partly determined by some other unmeasured factor P (such as private tutoring), then the actual situation might be that any case with XSp has a zero "chance" of getting a degree and any case with XSP has a 100% "chance". Similarly, in society B, dropout may not be random but caused by unmeasured factors. For these reasons we think that, in societies A and B, the use of Equation 1 is not so clearly justified as it is for society C. Rather, since it is likely *not* the case that all individuals with X=1 have an equal *random* chance of achieving the degree, we would seem instead to have:

Equation 3

{X=1} = > {**Proportion** of such cases with Degree) =0.9}.

Equation 4

{Having a degree} => {X=1}

We think that Lieberson, who tends to argue that we should assume "as if" probabilistic processes when we can't fully model the social world (1997a, especially pp. 364–5 and pp. 374–5), may not pay enough attention to such differences as that between societies C and A/B. Some of Ragin's references to "probabilities" in the context of QCA also seem open to the same line of criticism. The use of a probabilistic description like that in Equation 1 needs to be explicitly justified and this, we think, requires knowledge of the processes and mechanisms within black boxes, and not merely a study of the regularities generated by such processes. Of course, the rhetoric of QCA recommends "going back to the cases" and using theory to achieve this goal (see De Meur et al., 2009, on the "black box problem") but many published QCAs don't – in practice – seem to address these process issues in a manner that would allow us to decide between solutions like Equation 1 and Equation 3. We now turn to a more difficult scenario, one we think that current standard practice might well misinterpret.

4: An invented world with very low suf-consistency but perfect suf-coverage for an outcome

We now consider a more complex situation, more similar than the simple X-Y model in Table 1 to normal QCA practice in that it includes several causal conditions. Crucially, its analysis in terms of sufficiency generates an apparently unacceptably low proportion of cases with the outcome in all rows of the truth table. At first glance this suggests that an analysis of this dataset, where suf-consistency is "too low" (but suf-coverage is, as in Table 1, high), in terms of quasi-sufficiency is inappropriate, *but we argue that it is only with secure knowledge of processes and mechanisms that we can safely move from low consistencies with sufficiency to this negative conclusion*. For this reason, the dataset / causal structure we discuss here is, in our view, something like a critical case for QCA and related approaches. We draw on the previous section to suggest a way of treating it.

The new dataset includes these binary factors:

- "H" indicates completion of higher education.
- "C" indicates membership of the dominant social class (of two).
- "M" indicates male.

- "A" indicates having "high" ability.
- "E" indicates membership of one particular ethnic group (of two).
- "S" indicates having a secondary education diploma.

At various points we make use of some or all of these. We consider in particular the perspective of a scholar who has access to the set {C, A, M, E, H} and who undertakes a set theoretic analysis of H as a function of {C, A, M, E}. In doing this, we assume that s/he analyses merely the regularities[19] that exist between the input to and output from the black box shown in Figure 2. We again assume initially that our scholar does not know in any detail what causal processes / mechanisms operate in the black box but is simply assuming that all of C, A, M and E are causally relevant for H, and has access to a secondary dataset containing {C, A, M, E, H}.

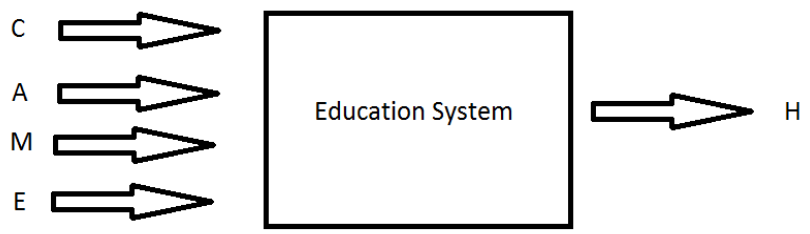

Figure 2: H=f(C,A,M,E)

Before considering these regularity-focused analyses, we describe the processes that we, *as inventors of this world,* know are operating within the black box, describe how these processes generate the observed regularities, and present a way in which these regularities can be described set theoretically which employs our knowledge of the nature of the generative processes. This description developed with knowledge of the underlying causal structure provides the background for our discussion of the work of our purely regularities-focused scholar. We, the inventors of this world, have arranged that the factors {C,A,M,E} completely determine, in a configurational manner, whether an individual achieves the important intermediate factor, S, a secondary school diploma[20]. This "S" is not, of course, one of the conditions available as part of the input/output model in Figure 2. We assume initially that the factor S is not available in the dataset to which our imagined scholar has access. The configurations leading to S are those set out in Table 5 while Table 6, columns 1-5, provides the full truth table for S=f{C,A,M,E}. We assume also that the number of cases in each row of Table 6 (the N_i) are identical[21].

Table 5: Configurations leading to S

c	A	M	e
c	A	M	E
C	A	m	e
C	A	m	E
C	a	M	E
C	A	M	e
C	A	M	E

Table 6: truth table for S=f{C,A,M,E}.

C	A	M	E	S	Proportion completing higher education (H=1)	Number
0	0	0	0	0	0	N_1
0	0	0	1	0	0	N_2
0	1	0	0	0	0	N_3
0	1	0	1	0	0	N_4
0	0	1	0	0	0	N_5
0	0	1	1	0	0	N_6
0	1	1	0	1	0.6	N_7
0	1	1	1	1	0.6	N_8
1	0	0	0	0	0	N_9
1	0	0	1	0	0	N_{10}
1	1	0	0	1	0.6	N_{11}
1	1	0	1	1	0.6	N_{12}
1	0	1	0	0	0	N_{13}
1	0	1	1	1	0.6	N_{14}
1	1	1	0	1	0.6	N_{15}
1	1	1	1	1	0.6	N_{16}

We, the inventors of this world, have also specified that all individuals achieving S wish to enter higher education (H). However, because of a scarcity of places, only a proportion of these can enter. Each configuration in Table 5, i.e. those with S=1, is taken *separately*, and all of its cases are entered into a lottery. Sixty percent of cases in each of these configurations enter higher education as a result of this random process. All of these then complete higher education. The resulting truth table is shown in the first six columns of Table 6. Before exploring in detail what QCA and CNA make of this dataset, we discuss briefly contrasting ways of understanding the regularities in Table 6. First, we can note that, on most views of what sufficiency is, none of the configurations approach being sufficient for H. The highest consistencies are 0.6. Secondly, we can also note that, for a researcher wanting to explain every individual case's actual outcome, any random process, like our lottery, makes this impossible. We cannot predict with certainty, on the basis of {C,A,M,E}, whether any individual belonging to the sets in Table 5 ends up with H. On the other hand, it is clearly necessary to be a member of one of the configurations in this table to achieve H. However, we can state, for any of these configurations, assuming that the lottery is truly random, that the probability of any individual case achieving H is 0.6. *Given, that is, that we, as the* *inventors of the data-generating processes, do have knowledge of the lottery-based generative mechanism*, we can use the form of description already used in Equation 1 and Equation 2 to represent these partly causal processes in a manner that does descriptive justice to the truth table. In non-minimised form, we obtain:

Equation 5

cAMe+cAME+CAme+CAmE+CaME+CAMe+CAME => {Probability(any such individual gets a degree)=0.6}

Equation 6

{Having a degree} => cAMe+cAME+CAme+CAmE+CaME+CAMe+CAME

Minimising, on a Quine-McCluskey basis, gives:

Equation 7

AC+AM+MCE => {Probability(any such individual gets a degree)=0.6}

Equation 8

{Having a degree} => AC+AM+MCE

Having set out this way of describing and analysing the truth table[22], *from the perspective of its inventors who know the processes inside the black box*, we now move to consider the set theoretic researcher who undertakes analyses without this knowledge of generative processes and mechanisms. His/her initial focus is again on the regularities. We do this in stages. First, employing fsQCA, we present and discuss a QCA of the simple input/output model in Figure 2. We then explore analyses using QCApro (Thiem, 2016a,b) and CNA (Ambuehl et al., 2014).

5: A QCA, employing fsQCA, of H=f(C,A,M,E)

What happens in practice, when a QCA focusing purely on regularities is performed employing the four inputs C, A, M and E and the output H? If fsQCA is used to assess the sufficiency of the various configurations of C, A, M and E then, using any consistency threshold recommended in the literature (usually above 0.75), no combination of these factors is found that is quasi-sufficient for H. So far, this represents an apparent failure to account for the outcome. A researcher might decide to break "best practice" rules to explore what happens when s/he sets a lower consistency threshold. S/he gradually lowers the threshold, testing H=f(C,A,M,E). Eventually, when a conventionally unacceptable (very low) threshold of 0.59 is reached, this parsimonious "solution" is found:

```
                 raw        unique
              coverage     coverage    consistency
             ----------   ----------   ----------
CA           0.571429     0.285714     0.600000
MA           0.571429     0.285714     0.600000
CME          0.285714     0.142857     0.600000

solution coverage: 1.000000
solution consistency: 0.600000
```

Now, while the overall solution and its configurational components do not reach levels that would allow any claims of quasi-sufficiency, the solution, according to a common interpretation of coverage, explains all of the outcomes[23]. However, any researcher who followed standard advice (Ragin, 2006; Schneider and Wagemann, 2012:133) only to assess coverage after quasi-sufficiency had been established would not, of course, get to this stage.

In QCA circles, a conventional way of describing this result, assuming it had not been dismissed because of the low suf-consistency scores, might be that being a member of the set AC+AM+MCE is necessary but is not sufficient for H. Our own view is that, *given knowledge of the lottery operation* – which would become available from process-tracing the careers of cases inside the black box – a better description is the one we used earlier: being a member of the set AC+AM+MCE is both necessary for H and sufficient to raise the probability of H for each case in these configurations to 0.6. Crucially, recommended practice concerning consistency levels has to be ignored to find this solution. In addition, we think that the description of the solution in terms of probability is only justified given secure knowledge that there is a random generative process operating here.

Since the above discussion suggests that our invented society is more easily open to a standard analysis of its regularities in terms of necessity than of sufficiency, it is worth looking at what the R-based package QCApro (Thiem, 2016a,b; see also Thiem and Duşa, 2013a,b for an earlier version) and Baumgartner's CNA (Ambuehl et al., 2014; Baumgartner, 2009, Baumgartner & Thiem, 2015a) – both of which take the analysis of necessity as seriously as that of sufficiency - make of these regularities. We start, next, with the R-based QCA, using the recently available version QCApro (Thiem, 2016a).

6: The R-based QCApro Package: addressing H=f(C,A,M,E)

First, sufficiency. If we use the eQMC function with a threshold of 0.75, QCApro reports no solution. However, if we drop the sufficiency threshold to 0.6, QCApro's eQMC function straightforwardly reproduces the fsQCA solution, and the output, through its use of the double implication sign and the reported coverage of 1.0 for the overall solution, also reports that "AC+AM+MCE" is necessary for H[24]. Similarly, employing the superSubset function for the model H=f(C,A,M,E) with a threshold of 0.75, we obtain, testing for sufficiency, no solutions. If we lower the threshold to 0.6, superSubset reports the components of the fsQCA solution for sufficiency, AC+AM+MCE. If we focus on necessity, using the superSubset function with a threshold of 0.75, the result is somewhat complicated to interpret:

		incl	cov.r
1	A	0.857	0.450
2	M+e	0.857	0.300
3	M+E	0.857	0.300
4	C+e	0.857	0.300
5	C+E	0.857	0.300
6	C+M	1.000	0.350
7	c+m+E	0.857	0.257
8	c+a+m+e	0.857	0.240

If instead we use a threshold of 1.0[25], we obtain the simpler:

		incl	cov.r
1	A+E	1.000	0.350
2	A+M	1.000	0.350
3	C+M	1.000	0.350
4	C+A	1.000	0.350

FsQCA reported AC+AM+MCE as necessary for H (with perfect nec-consistency). This superSubset result is clearly compatible with the fsQCA result. For example, consider M+C. Then, if both of M and C are not given, then none of the terms of "AC+AM+MCE" are satisfied[26]. We can further note from the superSubset results that (A+E)(A+M)(C+M)(C+A) is necessary for H. Multiplying this out and simplifying the result using the rules of Boolean algebra produces the result that AC+AM+EMC is necessary for H.

Summarising, a scholar who began with a QCApro analysis using eQMC and a threshold of 1.0 for suf-consistency but gradually lowered this, in the light of "no solutions", until 0.6 was reached, would eventually achieve the solution, AC+AM+MCE <=> H. The same would apply to a superSubset analysis of sufficiency. Undertaking a superSubset analysis of necessity, starting with a nec-consistency of 1.0 would produce a result from which H => AM+CA+CME can be derived. Crucially, as in the previous section, recommended practice concerning consistency thresholds has to be ignored to find the correct solution of the lottery-based dataset. In addition, to interpret correctly the low consistency / high coverage overall eQMC solution (see Equation 7, Equation 8) requires knowledge of the causal / generative processes operating here.

7: A CNA of H=f(C,A,M,E)

Initially following its creator's advice, we explore how CNA copes with our {C,A,M,E,H} dataset and, subsequently, since CNA has been designed to analyse chain structures (Baumgartner, 2009), we explore how it copes with the chain through S embedded in {C,A,M,E,S,H}, the dataset augmented by adding S.

Baumgartner (2015) argues that, in undertaking Boolean analyses of causes, we are seeking Boolean difference-makers, defined thus:

> Boolean difference-making (BD): A factor A is a Boolean difference-maker of an outcome E if, and only if, A is contained in a minimally sufficient condition AX of E such that AX, in turn, is contained in a minimally necessary condition of E.

Given the model H=f(C,A,M,E) then, in the solution "AC+AM+MCE <=> H", all of A, C, M and E are such difference-makers, as long as we interpret the => part of the equation in the probabilistic fashion described earlier[27]. Baumgartner and Epple (2014) also argue that, in analysing sufficiency, the suf-coverage should be given as much weight as the suff-consistency[28]:

> In the QCA literature, usually, only lowest bounds are provided for suf-consistency thresholds. For instance, Schneider and Wagemann (2010) recommend a lowest bound of 0.75 for suf-consistency. We contend, however, that there are good reasons to impose lowest bounds at least for suf-coverage of whole solution formulas as well. The suf-coverage of a solution formula being low means that it only accounts for few instances of an outcome. Or differently, in many cases where the outcome is given, there are causes at work that are not contained in the set of measured factors. However, unmeasured causes are likely to confound the data. The existence of potential confounders casts doubts on the causal interpretability of all other dependencies subsisting in the data, even on dependencies of perfectly consistent sufficiency. For uncontrolled causes might be covertly responsible for some of the dependencies manifest in the data. That is, the more likely it is that our data are confounded by uncontrolled causes, the less reliable a causal interpretation of resulting solution formulas becomes. In our view, suf-coverage of solution formulas should be used as a measure for the likelihood of confounding. The higher the coverage, the less likely it becomes that we are facing data confounding, the more reliable a causal interpretation of resulting solution formulas. We hence submit the same lowest bound for suf-coverage of solution formulas as usually imposed on suf-consistency: 0.75.

A coverage figure of 1 is the ideal, since all cases are then explained. However, it is noted in the R package manual[29] for CNA that:

> Note that the default consistency and coverage cut-offs of 1 frequently will not yield any atomic solution formulas because real-life data tend to feature noise due to uncontrolled background influences. In such cases, users should gradually lower consistency and coverage cut-offs (e.g. in steps of 0.05) until CNA finds solution formulas – for the aim of a CNA is to find solutions with the highest possible consistency and coverage scores. Consistency and coverage cut-offs should only be lowered below 0.75 with great caution. If cut-offs of 0.75 do not result in solutions, the corresponding data feature such a high degree of noise that there is a severe risk of causal fallacies.

Now, we don't have any "noise" in the sense referred to here in our black box, but rather a perfectly understandable lottery. We can expect, therefore, that using CNA to explore our dataset will raise some problems not covered by these quotes. We also already know that the model H=f(C,A,M,E) can produce a solution coverage of 1 but a solution consistency of only

0.6. Since the default setting for consistency thresholds in CNA is 1, it is no surprise that, if we request a default analysis of H as a function of C, A, M and E, then no solutions are reported. We next, therefore, gradually lower the suf-consistency and suf-coverage thresholds to explore solutions that are less perfect. If we initially lower them in tandem, following the advice in the first quote above, we obtain, once 0.6 is reached:

```
--- Coincidence Analysis (CNA) ---

Causal ordering:
A, M, C, E < H

Minimally sufficient conditions:
--------------------------------
Outcome H:
  condition consistency coverage
 A*C   -> H       0.600    0.571
 A*M   -> H       0.600    0.571
 C*E*M -> H       0.600    0.286

Atomic solution formulas:
-------------------------
Outcome H:
         condition consistency coverage
 A*C + A*M   <-> H       0.600    0.857
 A*C + C*E*M <-> H       0.600    0.714
 A*M + C*E*M <-> H       0.600    0.714

Complex solution formulas:
--------------------------
           condition consistency coverage
   (A*C + A*M <-> H)       0.600    0.857
 (A*C + C*E*M <-> H)       0.600    0.714
 (A*M + C*E*M <-> H)       0.600    0.714
```

CNA, because of our having given equal weight to suf-consistency and suf-coverage, has *not* delivered the solution "AC+AM+MCE <=> H" (suf-consistency=0.6, suf-coverage=1.0). Its components are present but have not been combined into the overall expression AC+AM+MCE. Faced with AC+AM+MCE , CNA finds it can remove a term and still meet the 0.6 consistency threshold for necessity (as well as that for sufficiency). The result is that the best solution, taking highest coverage as our guide, would be "AC+AM <-> H (with suf-consistency=0.6 and suf-coverage=0.857). We have lost CEM and, with it, perfect coverage. One way to try to avoid the outcome we have here is to rerun our looped CNAs but without the suf-consistency and suf-coverage being equated. Here, in ignoring the advice in the first quote above, we are perhaps, by aiming "to find solutions with the highest possible consistency and coverage scores", following one possible interpretation of the advice in the second. Setting suf-coverage to 1, for example, we can gradually reduce the threshold for suf-consistency from 1.0 to 0.55, recording the CNA output. Then suf-coverage can be set to 0.95, while suf-consistency reduces from 1.0, and so on. Running this double loop, starting with suf-coverage set to 1, we first encounter a solution when suf-consistency reaches 0.6. It is, now, as expected, "AC+AM+CEM <=> H" (suf-consistency=0.6, suf-coverage=1)[30]. Once again, as with fsQCA and QCApro, we have to revise the recommended practice to reach this solution. While the R CNA package manual correctly warns that the use of thresholds below 0.75 leads to a severe risk of causal fallacies we think that, in the context of datasets whose generation involves truly random processes, this warning might need qualification.

8: A CNA adding S, i.e. of H=f(C,A,M,E,S)

What about models including S? Figure 3 shows the chain we would like to see reported by CNA.

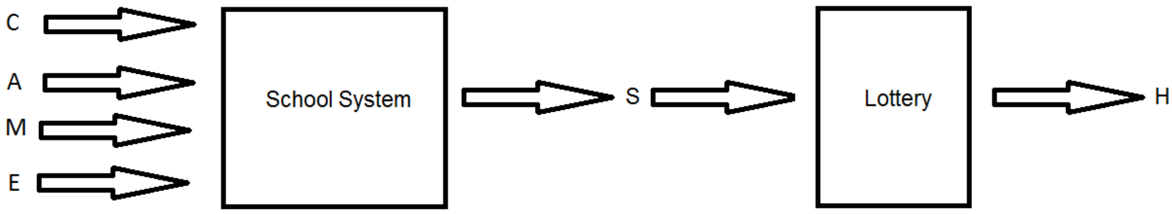

Figure 3: H=f(C,A,M,E,S)

What actually happens? First, employing the default settings of 1 for consistency and coverage, and the causal ordering given of A, M, C, E < S < H (which assumes some hypothetical knowledge of causal relations), we obtain:

```
--- Coincidence Analysis (CNA) ---

Causal ordering:
A, M, C, E < S < H

Minimally sufficient conditions:
--------------------------------
Outcome S:
  condition consistency coverage
 A*C   -> S       1.000    0.571
 A*M   -> S       1.000    0.571
 C*E*M -> S       1.000    0.286

Atomic solution formulas:
-------------------------
Outcome S:
               condition consistency coverage
 A*C + A*M + C*E*M <-> S       1.000    1.000

Complex solution formulas:
--------------------------
                 condition consistency coverage
 (A*C + A*M + C*E*M <-> S)       1.000    1.000
```

We have a perfect solution for S but none for H, since the default threshold for sufficiency is set too high to allow a solution for H. Running CNA again, reducing suf-consistency and suf-coverage *in tandem*, as advised in the earlier quote, we finally obtain, when suf-consistency=suf-coverage=0.6, these complex solutions:

```
Complex solution formulas:
--------------------------
                            condition consistency coverage
           (S <-> H)  *  (A <-> S)       0.600    0.857
   (A*C + A*M <-> H)  *  (A <-> S)       0.600    0.857
 (A*C + C*E*M <-> H)  *  (A <-> S)       0.600    0.714
 (A*M + C*E*M <-> H)  *  (A <-> S)       0.600    0.714
           (S <-> H)  *  (C <-> S)       0.600    0.714
   (A*C + A*M <-> H)  *  (C <-> S)       0.600    0.714
 (A*C + C*E*M <-> H)  *  (C <-> S)       0.600    0.714
 (A*M + C*E*M <-> H)  *  (C <-> S)       0.600    0.714
           (S <-> H)  *  (M <-> S)       0.600    0.714
   (A*C + A*M <-> H)  *  (M <-> S)       0.600    0.714
 (A*C + C*E*M <-> H)  *  (M <-> S)       0.600    0.714
 (A*M + C*E*M <-> H)  *  (M <-> S)       0.600    0.714
```

Solutions for H now appear and we have some chains such as, taking account of causal ordering, A -> S -> H. Once again, however, equalising suf-consistency and suf-coverage has produced less good solutions than we know, by design, are available. In particular, coverage is always less than one. If, using our knowledge of the causal/generative structure, we set suf-coverage to 1 and suf-consistency to 0.6, we obtain:

```
--- Coincidence Analysis (CNA) ---

Causal ordering:
A, M, C, E < S < H

Minimally sufficient conditions:
--------------------------------
Outcome H:
  condition consistency coverage
 S     -> H       0.600    1.000
 A*C   -> H       0.600    0.571
 A*M   -> H       0.600    0.571
 C*E*M -> H       0.600    0.286

Outcome S:
 condition consistency coverage
    A -> S       0.750    0.857
    C -> S       0.625    0.714
    M -> S       0.625    0.714

Atomic solution formulas:
-------------------------
Outcome H:
               condition consistency coverage
 A*C + A*M + C*E*M <-> H       0.600    1.000
 S                 <-> H       0.600    1.000

Complex solution formulas:
--------------------------
                 condition consistency coverage
 (A*C + A*M + C*E*M <-> H)       0.600    1.000
                 (S <-> H)       0.600    1.000
```

Now we have two solutions for H, with perfect coverage. However, we have not obtained the perfect solution we want for S, i.e. “AC+AM+CEM <-> S” (which would have allowed chaining) but rather “AC+AM+CEM <-> H”. The reason appears to be as follows. The use of a threshold low enough to allow a solution for H has allowed solutions like A->S to appear and therefore ruled out our obtaining AC+AM+CEM -> S, since AC, AM and CEM are not *minimally* sufficient at this consistency level. Consider, as an example, AC. As we know, AC is sufficient at a threshold of 1, while A is at 0.75, and C is at 0.625. Therefore removing either A or C from AC doesn't reduce the suf-consistency of the resulting terms below 0.6 and so CNA does remove them to achieve more minimally sufficient conditions. This leaves us with A->S, C->S and M->S. However, it happens that, as a result of the relations between the three sets A, C and M, that the suf-consistency of A+C+M falls below 0.6. It is 0.5, as an analysis using A+C+M as a combined factor shows. As a result, CNA doesn't report A+C+M=>S as a complex solution. These results arise from our use of one consistency level across the two parts of the chain in this integrated analysis of S and H. The solution would appear to be – given our knowledge, as designers of this dataset, of all the factors and their causal ordering – to undertake two separate analyses, using different consistency thresholds for S and H (1 and 0.6 respectively). The following analyses illustrate this. Consider S. If we analyse S=f(C,A,M,E), with a suf-consistency of 0.6, we obtain

```
--- Coincidence Analysis (CNA) ---

Causal ordering:
A, M, C, E < S

Minimally sufficient conditions:
--------------------------------
Outcome S:
 condition consistency coverage
    A -> S       0.750    0.857
    C -> S       0.625    0.714
    M -> S       0.625    0.714

Atomic solution formulas:
-------------------------
*none*

Complex solution formulas:
--------------------------
*none*
```

Here, since A+C+M => S only has a suf-con of 0.5, this analysis with a threshold of 0.6 doesn't report this disjunction of the minimally sufficient conditions A, C and M as a complex solution. Now compare this analysis, of S=f(C,A,M,E), but with suf-consistency set to 1:

```
--- Coincidence Analysis (CNA) ---

Causal ordering:
A, M, C, E < S

Minimally sufficient conditions:
--------------------------------
Outcome S:
  condition consistency coverage
 A*C   -> S       1.000    0.571
 A*M   -> S       1.000    0.571
 C*E*M -> S       1.000    0.286

Atomic solution formulas:
-------------------------
Outcome S:
               condition consistency coverage
 A*C + A*M + C*E*M <-> S       1.000    1.000

Complex solution formulas:
--------------------------
                 condition consistency coverage
 (A*C + A*M + C*E*M <-> S)       1.000    1.000
```

Here we have the correct solution for S. It seems then, to solve this {C,A,M,E}->S->H chain optimally, we need to run CNA twice[31], once for S=f(C,A,M,E) with suf-consistency=suf-coverage=1[32] and once for H=f(S) with suf-consistency=0.6, suf-coverage=1.

9: Discussion

We should stress that the work reported here is exploratory. We hope that others find it useful but that they will, where necessary, draw attention to gaps and/or errors. The crucial feature of our invented societies has been the insertion of an intermediating random generative process between a set of deterministic causes and an outcome, with a lottery standing in to simulate generative randomness. To clearly differentiate the latter from any randomness associated with measurement or sampling error, we also assumed that we had population level data and there was no measurement or calibration error. This has enabled us to explore ways in which the solutions generated by fsQCA, QCApro and CNA are affected by a set-up in which low suf-consistency is associated with high suf-coverage, but where there is a clear sense in which there is nothing to be done to improve further the suf-consistency parameter.

We first list what seem to the important findings and implication for current standard/best practice. We then reflect on these in a little more detail.

1. While it is usually recommended that suf-consistency thresholds should be set no lower than 0.75[33], and that suf-coverage should only be assessed where there is prior evidence of quasi-sufficiency, our analysis of the regularity relations characterising H=f{C,A,M,E} seems to provide a counterargument. We have shown that it is possible, given generative randomness, to have a full explanation of an outcome –

reflected in a solution suf-coverage of 1 – while having a solution suf-consistency well below the normally recommended threshold. Contrary to what is often said, low consistency thresholds do not have to indicate "noise" (in the sense of poor measurement and/or sampling) and/or fallacious causal claims. For this reason alone, scholars should give equal importance to sufficiency and necessity.

2. Related to this, we have suggested that our particular low suf-consistency / high suf-coverage situation can be described in a manner that draws on the usual account of probabilistic causation (where causes raise or lower the probability of outcomes) but incorporates Boolean ideas by stating that the conditions in the solution are sufficient to raise the probability of the outcome to the suf-consistency of the solution (see Equation 1 and Equation 2). We think this approach was already implicitly present in Ragin (1995).
3. Crucially, however, we argued that this move is only properly justified where the scholar has enough knowledge of the mechanism and processes linking the causal conditions and the outcome to be able to show that, *for each case*, there is a particular dispositional probability of achieving the outcome, given some particular values for the causal conditions in the model. In the absence of such knowledge, i.e. if the scholar is relying entirely on an analysis of regularities between conditions and outcome, the use of the term probability in our equations rather than proportion is not properly justified[34]. To move from proportion to probability, process-tracing and/or plausible theorising are needed in addition to regularities.
4. We have also shown that the use of CNA to uncover chains linking conditions to outcomes becomes less straightforward when one of the conditions in an analysis of regularities is actually a proxy for some random generative process. It seemed to be necessary to run two analyses using different consistency thresholds to find our invented chain through S from {C,A,M,E} to H.

Now, are these conclusions justified? How relevant are they to users of set theoretic methods? It might be argued, in response, that any scholar worth their salt would soon be able, via process-tracing, to find the lottery embedded in our invented world and then would move to deliver an analysis of S=f{C,A,M,E}with a solution suf-consistency and suf-coverage of 1 plus a note to explain that 60% of those entering the lottery (S) gained H. This is obviously so, but such a response would miss the point of our exploratory exercise. Although our use of a simple and transparent example[35] allows this response, our purpose has been to show that, if there indeed exist random generative processes of the sort that Lieberson and Goldthorpe argue characterise the social world (whether they be like our easy to uncover lottery or more like Cournot effects), then several things follow. First, a conventionally very low suf-consistency may be acceptable – we could have set our lottery to produce a very low solution suf-consistency of 0.3 – as long as it is associated with a high solution suf-coverage *plus* enough knowledge of causal mechanisms to justify the claim that this lowness is due to some source(s) of generative randomness rather than causal conditions having been omitted from the model. However, such combinations of low suf-consistency and high suf-coverage will only be found if a scholar undertakes analyses that ignore the often given advice to only assess suf-coverage once suf-consistency has been established at the 0.75 level. Recent work that stresses the equal importance of sufficiency and necessity in Boolean analyses of causation is helpful here (Baumgartner, 2015; Thiem, 2015a). Also helpful are the implementations of set theoretic methods in R (and, we should add, STATA, Longest and Vaisey, 2008) that allow the analyst, while analysing the regularities in a dataset, to use programming loops to run quickly and efficiently through a range of thresholds for consistency and coverage.

There are some issues that we have not addressed in what is already a long paper. We merely note one of these but say a little more about another. First, in the work reported here, we have bracketed out measurement and sampling error, whether random or non-random. Unfortunately, from the perspective of those who want easy conclusions, any scholar undertaking analysis of real as opposed to invented data is faced with these as well as any problems arising from random generative processes. Further work is needed to help those undertaking set theoretic analysis to identify the separate effects of these on consistency and coverage. Clues can be found, of course, from the solution parameters. Introducing measurement error or sampling error into our {C,A,M,E,S,H} dataset would, for example, except in exceptional circumstances, reduce the solution coverage below 1. However, alongside what might be learned from the solution parameters, all of (i) knowledge of generative mechanisms and processes, (ii) knowledge of the sampling and measurement procedures that produce any truth table, and (iii) careful theoretical reflection and judgement, will be required if a scholar is to make progress here.

Second, putting to one side the Laplacian claim that lotteries, dice-throwing, etc. are only apparently random devices, the question arises of whether our set-up is a causal structure or merely, at best, a partially causal structure, given that we cannot predict the exact outcome (H or not H) for all individual cases. Schneider and Wagemann (2012: 281) make an important relevant claim, arguing that solution formulae should not be the end of a QCA. Rather, the solutions and their parameters of fit should be "related back to the individual cases …

Researchers should make clear which cases – mentioned by their proper names – are (uniquely) covered by which of the paths in the solution formula (typical cases), and which are responsible for lower levels of consistency or coverage (deviant cases).” We agree that, in general, this is an important and worthwhile goal. However, given something like our lottery intervening in the paths to an outcome, the language here would perhaps need revising. In our analysis of H=F{C,A,M,E}, we found that AC+AM+MCE <=> H with suf-consistency of 0.6 and suf-coverage of 1. Now, some of the cases in, for example, the configuration AC will have gained H, some not. It’s not clear that the two categories are, in the face of the generative randomness operating here, appropriately described as respectively “typical” and “deviant”. Since if, taking the lottery into account, we look at S=f(C,A,M,E}, we have perfect suf-consistency and suf-coverage, i.e. no “deviant” cases, generative randomness does seem, once again, to result in a set-up that requires some rethinking of the advice that has been built up from thinking about deterministic models plus/minus some “noise” due to measurement error, sampling error or omitted factors[36].

Whether Boudon, Lieberson, Goldthorpe and others are correct in arguing that there are properly, or at least “as if”, random processes in the world that social researchers must take account of in their analyses will continue to be debated. However, we think we have shown that, for any set theoretic scholars who do accept their arguments, there are serious implications to consider. We hope our exploration of some of these proves helpful.

---

[1] There is a longstanding philosophical debate about “probability” in general and its relevance to explanation in particular disciplines. It is possible to take the position that, were we to have enough theoretical understanding, knowledge of initial conditions and computing power we might accurately predict the outcome of, for example, a die throw. This was Laplace’s position: see Hacking (1990: 11-12). As Nagel (1961) has noted, even the claim that there are essentially random processes operating at the quantum level may be revised. However, we assume here, with Lieberson (1997a,b) and Goldthorpe (2016), that chance and randomness are features of the social world, and that these can be simulated by such devices as die-throwing. For all practical purposes, we think, we can treat a six-sided die, thrown according to appropriate rules, as a set-up that simulates an “objective”, dispositional, view of probability (Popper, 1959).

[2] The scare quotes are deliberate here, since, given the history of philosophical debate about both the nature of probability and its relevance for social analysis, we need to take care in our own use of the phrase. We note one problem straightaway. Lieberson (1997a,b), in his discussion of “chance” in social life, has found inspiration and useful models in the analysis of sporting outcomes. An important question, but one Lieberson tends to avoid, is exactly when is it appropriate to treat a proportion or an average as reflecting randomness. He writes, for example (1997b), when discussing how rules can affect the relative importance of chance factors, “In figure skating competitions, suppose there is a difficult move that has a strong chance of failing - say on average 40% of the time competitive skaters fall when executing this step” (p.22). His subsequent discussion seems to assume

that such proportions as this 40% can be taken to represent probabilities. Presumably the justification for this position is (i) that there are chance factors operating here and (ii) in a run of attempts skaters have fallen 40% of the time. We take the view that considerable caution is required in making such moves from proportions ("relative frequencies") to "true" probabilities. We argue it is dangerous to rely too readily on "as if" notions of chance when faced with less than perfect accounts of some outcome, especially when analysis is based simply on partial regularities.

[3] Thiem et al. (2016) have noted that much work in this area has been limited by being tied to a particular study and dataset. They provide a more general perspective and alternative way forward.

[4] Marx (2006) does discuss randomness, but in the context of the claim by Lieberson (2004) that QCA cannot distinguish real from randomly generated data. We think that the distinction referred to, between "randomly assigned values (a data matrix with no meaning)" and a "table based on real data" picks out something other than what we discuss here. We assume that real data can derive from a partly random process, as when a lottery allocates college places.

[5] Schneider and Wagemann, in their textbook, (2012, chapter 11) refer to a variety of causes of less than perfect consistency and coverage but not, we think, to generative randomness. They also (316-7) argue that the use of fuzzy sets rather than crisp sets can be seen "as a defence against determinism". However, we think that our current argument supports the view that indeterminism can be handled by crisp and fuzzy sets. Thiem and Duşa's (2013b) textbook also, we think, doesn't discuss generative randomness.

[6] Little (1995), in a similar manner, discusses truth tables in terms of conditional probabilities, with probability defined as the frequency of cases with the outcome in a given subset (p. 9). He also employs the notion of "probabilistic causal relations", allowing a move away from a focus on "exceptionless regularities" to a perspective where particular combinations of conditions change the probability of an outcome (p.11). He then simulates explanations of revolution, making the weights of the exogenous factors themselves in his "possible worlds" probabilistic. Simultaneously, he makes the links between these factors, some intervening factors and an outcome probabilistic in each world. Whereas we explore the consequences of a random generative process in *one* otherwise deterministic world, Little uses his probabilistic modelling to construct as many as 700 different societal settings. Our cases are individuals, his are these settings. He uses an analysis of his simulated worlds – partly correlational, partly based on the relation between causal independence and conditional probabilities – to argue that theories of causal mechanisms are needed to make sense of the linkages in such datasets. We agree. More recently, Rohwer (2011) has discussed deterministic and stochastic functional models in relation to QCA. He argues, focusing on "uncertainty" in prediction, that, given what are usually termed "contradictions" in truth tables, it may be useful to consider stochastic functional models rather than to aim for compatibility with a deterministic model by data modification (by adding conditions, recalibrating, etc.). A consequence of using such stochastic models, he argues, is that it would no longer be possible to think of causes in terms of necessary and sufficient conditions (p. 738). We argue that, faced with a dataset that reflects generative randomness, rather than "uncertainty" of some unspecified type, it remains appropriate to use these terms.

[7] The "probabilities" Ragin refers to might in some cases be better described as *proportions* of cases in any configuration that achieve the outcome. An important question is when is it appropriate to treat such proportions as reflecting real "probabilistic processes" rather than, for example, an underspecified model. Feynman (1963),

arguing we speak of probability only when contemplating *future-oriented estimates*, wrote: "the probability of an outcome is our estimate for the most likely fraction of a number of repeated observations that will yield the outcome". However, there is an important difference between being able to claim that the estimated probability of throwing a six with a standard die is 1/6 because of the (dispositional) symmetry properties of the die and being able to claim it is 1/6 on the basis of an experiment with some particular die. When we are not treating die throws but rather the sort of data that Ragin is discussing, the latter of these two claims is likely to lead to model-based errors in estimating "probability" simply because we cannot always distinguish between there being random processes in the world (which might justify the claim that, for any *individual* with certain characteristics, the *probability* of the outcome is *p*) and our having an underspecified model of a non-probabilistic process (where we might more accurately say that, for the *set* of individuals with these characteristics, the *proportion* with the outcome is *p*). Sampling and measurement error will, of course, add difficulties when we want to claim that proportions reflect *generative* "probabilistic processes".

[8] Becker (1994) and Bandura (1982) discuss similar events. Rather than using "random", Becker uses the terms "chance" and "contingency". Bandura uses such terms as "chance encounters", "fortuitous encounters", "happenstance", "fortuitiveness", "chance", "fortuity" and "coincidence".

[9] Bandura (1982: 749) also notes that an event described as a "chance encounter" might be explained by reference to other causal narratives: "Although the separate chains of events in a chance encounter have their own causal determinants, their intersection occurs fortuitously rather than through deliberate plan (Nagel, 1961)."

[10] See Hacking (1990: 12).

[11] Goldthorpe (2016: 45-46), referring to "essential chance", uses this example but takes it from Monod's (1971) *Chance and Necessity*. Monod distinguishes "essential chance" from "operational chance", the latter referring to the need to take a probabilistic approach to phenomena which may actually be fully determined, but where it is practically impossible to employ a deterministic approach.

[12] The "thought to be" matters. See, for example, Nagel (1961: 335).

[13] Some might argue that, actually, the outcome of any die throw is fully determined. For all practical purposes, however, a die, if properly symmetric, if thrown according to agreed rules in some particular context, has the dispositional property that each side is equally likely to appear. Given six sides, we can then reason that the chance of any side appearing is 1 in 6.

[14] A lottery, in principle, can be rerun, as can the pulling of a handle on a casino machine. There are other cases of chance in social life where repeating the same event is less imaginable. In an election the result might be partly determined by variations in turn-out between constituencies due to weather that happen, contingently, to be correlated with varying party support across constituencies. This sort of contingency (Cournot-like) is therefore likely to be different to the sort of process our lottery represents. We choose the lottery route because it is easier to simulate and interpret.

[15] For a more general focus on regularities, see Pearl (2009).

[16] Researchers employing secondary datasets also are limited by those variables present.

[17] We should note that, were analyses to be undertaken of these three societies that included some or all of the missing intervening factors, then it would be possible, in some cases, to choose between complex, parsimonious or intermediate solution types (Ragin, 2008). In Society A, for example, an analysis using fsQCA of Y=f(X,S,Z) produces a complex solution Y=XSZ but a parsimonious one Y=Z. This regularities-based analysis of Society A would clearly force the analyst to ask further questions concerning both solution type and what counts as a correct "causal" account. While, in analyses of the conditions taken one at a time, Z is sufficient for Y with a consistency of 1, S with a consistency of 0.9 and X with a consistency of 0.9, a scholar might want to argue that the real cause of Y is "high" ability and, that were a measure of ability available that picked out higher levels of ability than X, then this revised X might have a consistency of 1 (Glaesser and Cooper, 2014). Here, however, we are using the discussion of intervening factors merely to illustrate and motivate our exploration of the consequences of randomness in a set theoretic context. Given that we only analyse the model Y=f(X) in this section, we won't expand this discussion.

[18] Williams & Dyer (2009), having contrasted, from a realist perspective, frequency, propensity and subjectivist interpretations of probability, argue that the propensity interpretation has "superior properties in respect of the explanation of the relationship of micro- or meso-level events to those at a macro-level" (85). They explore the use of cluster analysis to estimate nested "single-case probabilities" for "mentally disordered offenders". Our use of a lottery to generate a truth table creates a similar but simulated set-up that allows us to explore how QCA behaves when confronted by probabilistic generative processes.

[19] We noted earlier that many scholars in this area argue for process-tracing, going back to the cases, the use of theory, etc. as part of configurational causal analysis. We share this position but, once again for methodological reasons, we focus initially on regularities (see Cooper and Glaesser, 2012; Glaesser, 2015; Glaesser & Cooper, 2011).

[20] We simply assume that there are mechanisms operating that would provide any required causal links.

[21] In Cooper & Glaesser (2016) we explore why and how case weights matter for set theoretic analyses but, here, our solution coverage of 1 allows us to make this simplest assumption without, we think, causing our arguments any problems.

[22] It is important to note here that this "probability" interpretation might become invalid if, having failed to include some of the required causal conditions in our model, we were then to reduce the consistency threshold below 0.6 to obtain a solution with perfect coverage. For example, if we were to omit E, then we would need to reduce the consistency threshold to 0.3 to obtain the solution AC+AM+CM for H with a solution coverage of 1 and consistency of 0.525. Now, although 45% of the cases in the configuration CM achieve H, it is no longer true that each case in CM has a 0.45 probability in the strict sense of achieving H. On the other hand, were we to run this analysis with a threshold of 0.6 but omitting E, then we would obtain the solution AC+AM for H, with a reduced solution coverage of 0.857 but with a solution consistency of 0.6, allowing us to retain a truly probabilistic interpretation for cases in AC+AM.

[23] If the researcher included S in the model, using a 0.59 cut-off, there would be differences between fsQCA's standard solution types. The parsimonious solution is just S (con=0.6, cov=1), while the complex and intermediate solutions (with S,A,C set as positive for the outcome) both add S to all components of the solution for H=f(C,A,M,E), giving CAS+MAS+CMES => H.

[24] We include these assignments to eQMC: sol.type = "ps", row.dom = FALSE,   min.dis = FALSE (Thiem 2016b:23).

[25] We make use here of our knowledge of the data-generating process, but some argue anyway for a high threshold for necessity (Schneider and Wagemann, 2012:278).

[26] If we include S in superSubset's necessary tests, we obtain fsQCA's solution in row 1, where cov.r, of course, equals the suf-consistency of the solution reported by fsQCA for the simple model, H=f(S):

```
        incl   cov.r
--------------------
1  S    1.000  0.600
2  M+C  1.000  0.350
3  A+E  1.000  0.350
4  A+C  1.000  0.350
5  A+M  1.000  0.350
--------------------
```

[27] See Equation 7. Strictly, these terms are not sufficient for H, but for S.

[28] Thiem (2015b) has argued against this.

[29] See *Package 'cna'* (April 18, 2015), https://cran.r-project.org/web/packages/cna/cna.pdf, downloaded 29th July 2015.

[30] Care is needed in choosing between the solutions this double loop produces. For example, when cov is set to 0.85, and con to 0.6, we get the single complex solution "AC+AM <-> H". The other two solutions we have previously seen, "AC+CEM <-> H" and "AM+CEM <-> H" disappear, since their coverage (0.714) becomes too low.

[31] Responding to Baumgartner & Epple (2014), Thiem (2015b) shows that QCA can be used sequentially to analyse chains. He also takes a different position from Baumgarter & Epple (2014) concerning the relation between model specification and coverage thresholds (see Thiem, 2015b:6). QCApro makes it easy to analyse several outcomes simultaneously. Using our dataset with all of {C,A,M,E,S,H} present in the analysis, the R code for QCApro below (where, note, we do need to employ an unconventional threshold of 0.6) will produce the elements of the chain amongst its listed outputs. In detail: the code "lottery.cna.QCApro <-eQMC(lotterymodeldataCAMESH, outcome = names(lotterymodeldataCAMESH), relation = "sufnec", incl.cut1 = .6, row.dom=FALSE, min.dis=FALSE)", following the author's advice on setting both row.dom and min-dis to FALSE for causal analysis (see also Baumgartner & Thiem, 2015c), produces:

```
There is no solution for outcome "C".
There is no solution for outcome "M".
M1: cS + eS + mS <=> A
There is no solution for outcome "E".
M1: CA + MA + CME <=> S
M1: S <=> H
M2: CA + MA + CME <=> H
```

[32] We've used our designer's knowledge here. In fact, using a threshold of 0.75 or above will produce the correct result. A loop could be used to explore results as the threshold is varied.

[33] See, for example, Schneider & Wagemann (2012: 127-8). Having noted there shouldn't be any hard and fast rule, they nevertheless note that values below 0.75 are often problematic. They rule out 0.5 for "obvious reasons" as then "half of the empirical evidence contradicts the subset relational statement of sufficiency". However, in the case of our lottery-based world, if we were to reset our winners' proportion to 50%, we could argue for accepting a threshold as low as 0.5. The authors do note that the specific research context matters. We seem to have invented a partially deterministic context that makes 0.6 (or even 0.5) an acceptable threshold.

[34] This problem is not one merely for set theoretic analyses. It applies to regression modelling and other forms of analysis.

[35] Other cases with similarities to our lottery-based simulation could be considered. For example, it is neither difficult or, arguably, unrealistic to imagine a world in which entry to a career demands certain qualification whose gaining by individuals might be largely causally accounted for, but where subsequent progress in the career is highly unpredictable because of Cournot effects.

[36] The literature on residuals analysis in regression modelling is likely to provide some help here.

References

Ambuehl, M., Baumgartner, M., Epple, R., Kauffmann, A. & Thiem, A. (2014): *cna: A Package for Coincidence Analysis*. R Package Version 1.0-3. URL: http://cran.r-project.org/package=cna.

Bandura, A. (1982). The psychology of chance encounters and life paths. *American Psychologist*, 37, (7), 747–755.

Baumgartner, M. (2008): Regularity theories reassessed. *Philosophia*, 36, 327–354.

Baumgartner, M. (2009): Inferring causal complexity. *Sociological Methods & Research*, 38, (1), 71-101.

Baumgartner, M. (2015): Parsimony and causality. *Quality and Quantity*, 49, (2), 839-856.

Baumgartner, M. & Epple, R. (2014): A Coincidence Analysis of a causal chain: the Swiss minaret vote. *Sociological Methods & Research*, 43, (2), 280-312.

Baumgartner, M. & Thiem, A. (2015a): Identifying complex causal dependencies in configurational data with coincidence analysis. *The R Journal*, 7, (1), 176-184.

Baumgartner & Thiem (2015b) "Often Trusted But Never (Properly) Tested: Evaluating Qualitative Comparative Analysis". https://www.researchgate.net/publication/280555425_Often_Trusted_But_Never_Properly_Tested_Evaluating_Qualitative_Comparative_Analysis. Downloaded 18/4/16.

Baumgartner, M., & Thiem, A. (2015c): Model Ambiguities in Configurational Comparative Research. *Sociological Methods & Research*. DOI: 10.1177/0049124115610351.

Becker, H.S. (1994). 'Foi por acaso': conceptualizing coincidence. *The Sociological Quarterly*, 35, (2), 183–194.

Boudon, R. (1986): *Theories of Social Change*. Cambridge: Polity Press.

Collier, D. (2011): Understanding process tracing. *PS: Political Science and Politics*, 44, 823-830.

Cooper, B. & Glaesser, J. (2011): Using case-based approaches to analyse large datasets: a comparison of Ragin's fsQCA and fuzzy cluster analysis. *International Journal of Social Research Methodology*, 14, (1), 31-48.

Cooper, B. & Glaesser, J. (2012): Qualitative work and the testing and development of theory: lessons from a study combining cross-case and within-case analysis via Ragin's QCA. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 13, (2), Art. 4.

Cooper, B. & Glaesser, J. (2015a): Exploring the robustness of set theoretic findings from a large n fsQCA: An illustration from the sociology of education. *International Journal of Social Research Methodology*, DOI: 10.1080/13645579.2015.1033799

Cooper, B. & Glaesser, J. (2015b): Qualitative Comparative Analysis, Necessary Conditions, and Limited Diversity: Some Problematic Consequences of Schneider and Wagemann's Enhanced Standard Analysis. *Field Methods*, doi:10.1177/1525822X15598974

Cooper, B. & Glaesser, J. (2016): Analysing necessity and sufficiency with Qualitative Comparative Analysis: how do results vary as case weights change? *Quality & Quantity*, 50, (1), 327-346.

De Meur, G., Rihoux, B. & Yamasaki, S. (2009) Addressing the critiques of QCA, in Rhoux, B. & Ragin, C.C. (Eds) Configurational Comparative Methods, Sage.

Feynman, R. (1963): *Feynman Lectures on Physics, Vol. I: 1*. Reading, Mass.: Addison-Wesley.

George, A.L. & Bennett, A. (2005): *Case studies and theory development in the social sciences*. Cambridge, Massachusetts: MIT Press.

Glaesser, J. (2015): *Young people's educational careers in England and Germany. Integrating survey and interview analysis via Qualitative Comparative Analysis*. Basingstoke: Palgrave Macmillan.

Glaesser, J. & Cooper, B. (2011). Selecting cases for in-depth study from a survey dataset: an application of Ragin's configurational methods. *Methodological Innovations Online* 6(2): 52-70.

Glaesser, J. & Cooper, B. (2014): Exploring the consequences of a recalibration of causal conditions when assessing sufficiency with fuzzy set QCA. International Journal of Social Research Methodology, 17, (4), 387-401.

Goldthorpe, J.H. (2007): *On Sociology. Second Edition. Volume One: Critique and Program.* Stanford: Stanford University Press.

Goldthorpe, J. (2016): *Sociology as a Population Science*. Cambridge: Cambridge University Press.

Hacking, I. (1990): *The Taming of Chance*. Cambridge: Cambridge University Press.

Lieberson, S. (1991): Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases. *Social Forces*, 70, (2), 307-320.

Lieberson, S. (1997a). The big broad issues in society and social history. Application of a probabilistic perspective. In: McKim, Vaughn R. and Turner, Stephen P. (eds.), *Causality in Crisis? Statistical methods and the search for causal knowledge in the social sciences*. Notre Dame, Indiana: University of Notre Dame Press, 359–385.

Lieberson, S. (1997b): Modeling social processes: some lessons from sports. *Sociological Forum*, 12, 11-35.

Lieberson, S. (2001): Review Essay on "Fuzzy Set Social Science" by Charles C. Ragin. *Contemporary Sociology*, 30, (4), 331-334.

Lieberson, S. (2004): Comments on the Use and Utility of QCA. *Qualitative Methods.* 2, (2), 13-14.

Little, D. (1995): An experiment in causal reasoning. Paper presented at the Spindel Conference in Philosophy, October 1995. URL: http://www-personal.umd.umich.edu/~delittle/causexp.PDF.

Longest, K.C. & Vaisey, S. (2008): fuzzy: A program for Performing Qualitative Comparative Analyses (QCA) in Stata. *Stata Journal* 8 (1):79-104.

Marx, A. (2006): Towards more robust model specification in QCA: results from a methodological experiment. *COMPASSS Working Paper* 2006-43. URL: http://www.compasss.org/wpseries/Marx2006.pdf

Monod, J. (1971): *Chance and necessity*. Collins: Glasgow.

Nagel, E. (1961): *The structure of science*. London: Routledge & Kegan Paul.

Pearl, J. (2009): *Causality: Models, Reasoning, and Inference*, 2nd Edition. Cambridge: Cambridge University Press.

Popper, K.R. (1959): The propensity interpretation of probability. *The British Journal for the Philosophy of Science* 10, (37), 25-42.

Ragin, C.C. (1987): *The Comparative Method. Moving beyond Qualitative and Quantitative Strategies.* Berkeley, Los Angeles, London: University of California Press.

Ragin, C.C. (1995): Using Qualitative Comparative Analysis to study configurations. In: Kelle, Udo (ed.), *Computer-aided qualitative data analysis. Theory, methods and practice*. London: Sage, 177-189.

Ragin, C.C. (2000): *Fuzzy-Set Social Science*. Chicago and London: University of Chicago Press.

Ragin, C.C. (2006): Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis*, 14, (3), 291-310.

Ragin, C.C. (2008): *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago Press.

Ragin, C.C., Mayer, S.E. & Drass, K.A. (1984): Assessing Discrimination: a Boolean Approach. *American Sociological Review*, 49, (April), 221-234.

Ragin, C.C. & Bradshaw, Y.W. (1991): Statistical analysis of employment discrimination: a review and critique. *Research in Social Stratification and Mobility*, 10, 199-228.

Ragin, C. C. and Davey, S. (2014). fs/QCA: Fuzzy-Set/Qualitative Comparative Analysis, Version 2.5 [Computer Program]. Irvine: Department of Sociology, University of California.

Rohwer, G. (2011): Qualitative Comparative Analysis: a discussion of interpretations. *European Sociological Review*. 27, 6, 728-740.

Schneider, C. Q. and Wagemann, C. (2010): Standards of Good Practice in Qualitative Comparative Analysis (QCA) and Fuzzy-Sets. *Comparative Sociology,* 9: 397-418.

Schneider, C.Q. & Wagemann, C. (2012): *Set-theoretic methods for the social sciences: A guide to Qualitative Comparative Analysis*. Cambridge: Cambridge University Press.

Schneider, C.Q. & Wagemann, C. (2015): Assessing ESA on what it is designed for: a reply to Cooper and Glaesser. *Field Methods*, doi:10.1177/1525822X15598977.

Skaaning, S.E. (2011): Assessing the robustness of crisp-set and fuzzy-set QCA results. *Sociological Methods & Research*, 40(2) 391-408.

Thiem, A. (2015a): Standards of good practice and the methodology of necessary conditions in Qualitative Comparative Analysis: a critical view on Schneider and Wagemann's Theory-Guided/Enhanced Standard Analysis. *COMPASSS WP Series* 2015-83, URL: http://www.compasss.org/wpseries/Thiem2015.pdf.

Thiem, A. (2015b): Using Qualitative Comparative Analysis for identifying causal chains in configurational data: a methodological commentary on Baumgartner and Epple (2014). *Sociological Methods & Research*, 44, (4), 723-736.

Thiem, A. (2016a): *QCApro: Professional Functionality for Performing and Evaluating Qualitative Comparative Analysis*. R Package Version 1.0-0. URL: http://cran.r-project.org/package=QCApro.

Thiem, A. (2016b): QCApro manual. https://cran.r-project.org/web/packages/QCApro/QCApro.pdf.

Thiem, A. & Duşa, A. (2013a): QCA: A Package for Qualitative Comparative Analysis. *The R Journal* 5 (1), 87-97.

Thiem, A. & Duşa, A (2013b): *Qualitative Comparative Analysis with R: A User's Guide*. New York: Springer.

Thiem, A., Spöhel, R. & Duşa, A. (2016): Enhancing sensitivity diagnostics for Qualitative Comparative Analysis: a combinatorial approach. *Political Analysis*, 24:104–120.

Williams, M. & Dyer, W. (2009): Single-case probabilities. In: Byrne, D. & Ragin, C.C. (eds.) *The Sage Handbook of Case-based Methods*. London: Sage, 84-100.