Causal Modeling with Multi-Value and Fuzzy-Set Coincidence Analysis

Michael Baumgartner* and Mathias Ambühl†

Abstract

Coincidence Analysis (CNA) is a configurational comparative method of causal data analysis that is related to *Qualitative Comparative Analysis* (QCA) but, contrary to the latter, is custom-built for analyzing causal structures with multiple outcomes. So far, however, CNA has only been capable of processing dichotomous variables, which greatly limited its scope of applicability. This paper generalizes CNA for multi-value variables as well as continuous variables whose values are interpreted as membership scores in fuzzy sets. This generalization comes with a major adaptation of CNA's algorithmic protocol, which, in an extended series of benchmark tests, is shown to give CNA an edge over QCA not only with respect to multi-outcome structures but also with respect to the analysis of non-ideal data stemming from single-outcome structures. The inferential power of multi-value and fuzzy-set CNA is made available to end users in the newest version of the R-package cna.

Keywords: configurational comparative methods, set-theoretic methods, Coincidence Analysis, Qualitative Comparative Analysis, Boolean causation

1 Introduction

Since the mid-1980ies, different variants of *configurational comparative methods* (CCMs), or, as they are also called, *set-theoretic methods*, have gradually been added to the toolkit for causal data analysis in the social sciences. CCMs are designed to investigate different hypotheses and uncover different properties of causal structures than traditional regression analytical methods (RAMs) and, thus, complement the latter (rather than compete with

*University of Bergen, Norway, michael.baumgartner@uib.no
†ConsultAG, Switzerland, mathias.ambuehl@consultag.ch

them) (Thiem et al. 2016). RAMs examine covariation hypotheses as "the more/less of X, the more/less of Y" that link exogenous and endogenous *variables*, and they quantify net-effects, that is, marginal effects in different causal backgrounds. By contrast, CCMs study implication hypotheses as "$X = \chi_i$ is sufficient/necessary for $Y = \gamma_i$" that link *specific values* of exogenous and endogenous variables. Moreover, instead of quantifying net-effects, CCMs lay a Boolean ordering over sets of causes by locating their elements on the same or different causal paths to the ultimate outcome. In other words, while RAMs investigate the quantitative properties of causal structures as characterized by statistical or probabilistic theories of causation (Simon 1954; Suppes 1970), CCMs scrutinize their Boolean properties as described by regularity theories of causation (Mackie 1974).

The Boolean properties of causation encompass three complexity dimensions. The first is *conjunctivity*: to bring about an effect, say, liberal democracy in early modern Europe ($D = 1$), different factors need to be instantiated (or *not* instantiated) jointly; for instance, according to Downing's (1992) theory of the origins of liberal democracy, a country must have a history of medieval constitutionalism ($C = 1$) and absent military revolutions ($R = 0$) (cf. Goertz 2006, 252-254). Only a coincident instantiation of the conjunction $C{*}r$ is sufficient to produce the effect D, where upper- and lowercase letters symbolize presence and absence, respectively. *Disjunctivity* is a second complexity dimension: an effect can be brought about along several alternative causal paths. Downing (1992, 78-79, 240) identifies four alternative paths leading to the absence of military revolution (r): a geography that deters invading armies (G), commercial wealth (W), foreign resource mobilization (M), and foreign alliances (A). Each condition in the disjunction $G + W + M + A$ can bring about the effect r independently of the other conditions. The third complexity dimension is *sequentiality*: effects tend to cause further effects, meaning that causal influence can be propagated along causal chains. In Downing's theory of liberal democracy there are multiple causal chains, for instance, W is causally relevant to r, which, in turn, is causally relevant to D, or there is a chain from A via r to D. Overall, the theory entails the following Boolean causal model for the development of liberal democracies (cf. Goertz 2006, 254), where "$\rightarrow$" stands for the Boolean operation of implication:

$$(G \ + \ W \ + \ M \ + \ A \ \rightarrow \ r) * (C{*}r \ \rightarrow \ D) \tag{1}$$

The most prominent CCM designed to infer Boolean models from data is *Qualitative Comparative Analysis* (QCA; Ragin 2008). Since its first introduction in Ragin (1987), QCA has gained considerable popularity and has been applied in areas as diverse as busi-

ness administration, management, environmental science, evaluation science, international relations, political science, public health, and sociology. Whereas the original variant of QCA—*crisp-set* QCA (csQCA)—is restricted to modeling the interplay among dichotomous variables, there meanwhile exist fully worked out variants that can process multi-value variables, *multi-value* QCA (mvQCA) (Cronqvist and Berg-Schlosser 2009), as well as continuous variables whose values are interpreted as membership scores in fuzzy sets, *fuzzy-set* QCA (fsQCA) (Ragin 2009).

However, all variants of QCA focus on the complexity dimensions of conjunctivity and disjunctivity only, as QCA always treats exactly one factor Z as endogenous and all other elements in an analyzed set of factors as exogenous, meaning as mutually independent potential direct causes of Z. QCA will thus not find a chain model as (1).[1] In light of this restriction, Baumgartner (2009a, 2009b) introduced a new configurational comparative method called *Coincidence Analysis* (CNA). As a member of the family of CCMs, CNA—just like QCA—investigates implication hypotheses and scrutinizes the Boolean properties of causation. Contrary to QCA, however, CNA searches for direct and indirect causal dependencies among all analyzed factors and is, hence, capable of uncovering all Boolean complexity dimensions: conjunctivity, disjunctivity *and* sequentiality. That is, CNA is tailor-made to recover chain models as (1).

So far, though, CNA has only been available in a crisp-set variant (csCNA) that requires dichotomous variables. As real-life data very often feature variables that can take more than two or even continuous values, the applicability of CNA has, hence, been greatly limited. This paper removes that limitation by generalizing CNA for multi-value variables (mvCNA) as well as for continuous variables whose values are interpreted as membership scores in fuzzy sets (fsCNA). This generalization comes with a major adaptation of the basic algorithmic protocol on the basis of which CNA builds causal models. In a nutshell, while CNA so far—just like QCA—adopted a *top-down approach* to model building that first identifies complete sufficient and necessary conditions of outcomes and then gradually eliminates redundant elements, the generalized variant of CNA uses a *bottom-up approach* that progressively combines factor values to complex but redundancy-free sufficient and necessary conditions.

The generalized CNA algorithm presented here has been implemented in a new version of the R-package cna, version 2.0, which makes the whole inferential power of mvCNA and

[1] In a recent comment on CNA, Thiem (2015) argues that QCA is not necessarily tied to an algorithm that is restricted to single-outcome structures. Thiem then suggests a QCA approach to searching for chains that much resembles CNA.

fsCNA available to end-users. By drawing on this software package and the currently most advanced R-package for QCA, QCApro (Thiem 2016), the paper also performs a whole battery of benchmark tests that evaluate and compare the performance of CNA and QCA when applied to data with varying forms of data deficiencies. The test series reveals that the reversal of the basic model building approach gives CNA an edge over QCA not only with respect to multi-outcome structures but also with respect to the analysis of non-ideal data stemming from single-outcome structures.

The paper is organized as follows. Section 2 introduces the theoretical background of CNA, its main input parameters, and exhibits the inferential potential and limitations of CNA. The generalization of the CNA algorithm is presented in section 3. Section 4 reports the results of the test series evaluating and comparing CNA and QCA. A replication script detailing all analytical steps is provided in the appendix.

2 Theoretical background

2.1 Boolean difference-making

As all CCMs, CNA searches for causal dependencies as defined by so-called *regularity theories* of causation, whose development dates back to Hume (1748) and (Mill 1843). Modern regularity theories spell out causation in terms of Boolean difference-making within a fixed context of causal background conditions. Somewhat more explicitly, X is a regularity theoretic cause of Y if there exists a (fixed) configuration of background conditions $\mathcal{F}$ such that a change in the value of X is regularly associated with a change in Y in $\mathcal{F}$. If X does not make a difference to Y in any context $\mathcal{F}$, X is redundant to account for Y and, thus, no cause of Y (Mackie 1974; Graßhoff and May 2001).

To render that idea more precise, some conceptual preliminaries are required. First, the relata of (i.e. the entities related by) regularity theoretic causation are *variables taking specific values* as $X = \chi_i$ or $Y = \gamma_i$; interchangeably, we will also speak of *factors* and their *values*. In case of csCNA and fsCNA, variables can be interpreted as sets and their values as membership scores in those sets. As is conventional, we shall abbreviate membership in a set X by upper case and non-membership by lower case Roman letters. An alternative interpretation, which lends itself particularly well for causal modeling, is that "X" stands for the presence of the factor X and "x" for its absence. In case of mvCNA, by contrast, we will not abbreviate value assignments to variables and, instead, use the explicit 'Variable=value' notation by writing, say, "$X = 3$" for X taking the value 3.

Apart from the Boolean operations of conjunction, disjunction, and negation, whose classical definitions are presupposed here, the implication operator "$\rightarrow$" and the equivalence operator "$\leftrightarrow$" are of core importance for the regularity theoretic definition of causation. According to a classical interpretation of the implication operator, an expression as "$X = 3 \ \rightarrow \ Y = 4$" states that whenever X takes the value 3, Y takes 4; or "$X \ \rightarrow \ Y$" states that whenever X is present, Y is present. These claims are true if, and only if (iff), there is no case satisfying the expression on the left-hand side of "$\rightarrow$" and not satisfying the expression on the right-hand side. Furthermore, "$X = 3 \ \leftrightarrow \ Y = 4$" and "$X \ \leftrightarrow \ Y$" are true iff the implication holds both ways, meaning that all cases satisfying the left-hand side of "$\leftrightarrow$" also satisfy the right-hand side, and vice versa.

For the subsequent generalization of CNA for variables with continuous values from the interval $[0, 1]$ the classical Boolean operations must be translated into fuzzy logic. There exist numerous systems of fuzzy logic (for an overview cf. Hájek 1998), each of which comes with its own rendering of Boolean operations. We will adopt the following fuzzy-logic renderings, which have become standard in the context of CCMs: conjunction $X{*}Y$ is defined in terms of the minimum membership score in X and Y, i.e. $\min(X, Y)$, disjunction $X + Y$ in terms of the maximum membership score in X and Y, i.e. $\max(X, Y)$, negation x in terms of $1 - X$, and implication $X \rightarrow Y$ in terms of a binary operation that takes the value 1 if the membership score of X is smaller or equal to Y ($X \leq Y$) and the value 0 otherwise.

The implication operator allows for defining the notions of sufficiency and necessity which figure at the heart of regularity theories. The presence of X is *sufficient* for the presence of Y iff $X \ \rightarrow \ Y$ holds; and the presence of X is *necessary* for the presence of Y iff $Y \ \rightarrow \ X$ holds. Analogously, the more complex expression $X = 3 \ + \ Z = 2$ is sufficient and necessary for $Y = 4$ iff $X = 3 + Z = 2 \ \leftrightarrow \ Y = 4$ holds.

Relations of sufficiency and necessity merely amount to patterns of co-occurrence of factor values; as such, they carry no causal connotations whatsoever. In fact, most dependencies of sufficiency and necessity do not reflect causal dependencies. For that reason, regularity theories rely on a *non-redundancy principle* as an additional constraint to filter out those relations of sufficiency and necessity that are due to underlying causal dependencies and, hence, are amenable to a causal interpretation. Causal structures do not feature redundant elements, meaning causes are those elements of sufficient and necessary conditions for which at least one configuration of background conditions $\mathcal{F}$ exists in which they are indispensable to account for a scrutinized outcome. In other words, whatever can be removed from sufficient and necessary conditions without affecting the latter's sufficiency and necessity is redundant

and, therefore, not causally interpretable. Only sufficient and necessary conditions that are completely free of redundant elements reflect causation (Baumgartner 2014).

2.2 Boolean causal models

Modern regularity theories formally cash this idea out on the basis of the notion of a minimal theory. There are atomic and complex minimal theories. An *atomic minimal theory* of an outcome Y is a minimally necessary disjunction of minimally sufficient conditions of Y (Graßhoff and May 2001). A conjunction Φ of coincidently instantiated factor values (e.g. $X_1{*}X_2{*}\ldots{*}X_n$) is a minimally sufficient condition of Y iff Φ is sufficient for Y ($\Phi \rightarrow Y$), and there does not exist a proper part Φ' of Φ such that $\Phi' \rightarrow Y$. A proper part Φ' of Φ is the result of eliminating one or more conjuncts from Φ. A disjunction Ψ of minimally sufficient conditions (e.g. $\Phi_1 + \Phi_2 + \ldots + \Phi_n$) is a minimally necessary condition of Y iff Ψ is necessary for Y ($Y \rightarrow \Psi$), and there does not exist a proper part Ψ' of Ψ such that $Y \rightarrow \Psi'$. A proper part Ψ' of Ψ is the result of eliminating one or more disjuncts from Ψ. Overall, an atomic minimal theory of Y states an equivalence of the form $\Psi \leftrightarrow Y$ (where Ψ is an expression in disjunctive normal form[2] and Y is a single factor value). Atomic minimal theories can be conjunctively concatenated to form *complex minimal theories*.

Minimal theories represent the causally interpretable dependencies of sufficiency and necessity among the factors in an analyzed data set. Or differently, minimal theories connect Boolean dependencies, which—by themselves—are purely functional and non-causal, to causal dependencies: those, and only those, Boolean dependencies that appear in minimal theories stem from underlying causal dependencies. Atomic minimal theories stand for causal structures with one outcome, complex theories represent multi-outcome structures.

To further clarify the causal interpretation of minimal theories, consider the following complex exemplar:

$$(A{*}b \; + \; a{*}B \; \leftrightarrow \; C) \, * \, (C{*}f \; + \; D \; \leftrightarrow \; E) \tag{2}$$

From a functional perspective, (2) claims that the presence of A in conjunction with the absence of B (i.e. b) as well as a in conjunction with B are two alternative minimally sufficient conditions of C, and that $C{*}f$ and D are two alternative minimally sufficient conditions of E. Moreover, both $A{*}b + a{*}B$ and $C{*}f + D$ are claimed to be minimally necessary for C and E, respectively. Against the background of a regularity theory of causation, these func-

[2]A Boolean expression is said to be in disjunctive normal form iff it is a disjunction of one or more conjunctions of one or more literals.

tional relations entail the following causal claims: (i) the factor values listed on the left-hand sides of "$\leftrightarrow$" are directly causally relevant for the factor values on the right-hand sides; (ii) A and b are located on the same causal path to C, which differs from the path on which a and B are located, and C and f are located on the same path to E, which differs from D's path; (iii) $A{*}b$ and $a{*}B$ are two alternative indirect causes of E whose influence is mediated on a causal chain via C. More generally put, minimal theories ascribe causal relevance to their constitutive factor values, place them on the same or different paths to the outcomes, and distinguish between direct and indirect causal relevancies. That is, they render transparent the three Boolean complexity dimensions of causality—which is why we shall likewise refer to minimal theories as *Boolean causal models*.

Two fundamentals of the interpretation of Boolean causal models should be reiterated (not least because the literature on CCMs is often not sufficiently clear about these points). First, ordinary Boolean models make claims about causal relevance but not about causal irrelevance. According to a regularity theoretic definition of causation, X_1 is a cause of an outcome Y iff there exists a context of background conditions $X_2{*}\ldots{*}X_n$ in which X_1 makes a difference to Y—meaning that $X_1{*}X_2{*}\ldots{*}X_n$ and $x_1{*}X_2{*}\ldots{*}X_n$ are associated with different Y-values (Mackie 1974, 59-87). While establishing causal relevance merely requires demonstrating the existence of at least one such difference-making context, establishing causal irrelevance would require demonstrating the non-existence of such a context, which is impossible on the basis of the non-exhaustive data samples that are typically analyzed in observational studies. Correspondingly, the fact that, say, a further factor value G does not figure in (2) does not imply G to be causally irrelevant to either C or E. The non-inclusion of G simply means that the data from which model (2) has been derived do not contain evidence for the causal relevance of G to C or E. However, future research having access to additional data might reveal the existence of a difference-making context for G and, hence, entail the causal relevance of G to C or E after all.

Second, Boolean models are to be interpreted relative to the data set δ from which they have been derived. They do not purport to reveal all of an underlying causal structure's Boolean properties but only detail those causally relevant factor values along with those conjunctive, disjunctive, and sequential groupings for which δ contains evidence. By extension, two different Boolean models $\mathbf{m}_i$ and $\mathbf{m}_j$ derived from two different data sets δ_i and δ_j are in no disagreement if the causal claims entailed by $\mathbf{m}_i$ and $\mathbf{m}_j$ stand in a subset relation. For example, model (3) and model (4) do not conflict with model (2):

$$(A \ + \ B \ \leftrightarrow \ C) * (C \ + \ D \ \leftrightarrow \ E) \tag{3}$$

$$(A{*}b \ + \ a{*}B \ \leftrightarrow \ C) * (C{*}f \ + \ D{*}G \ \leftrightarrow \ E) \tag{4}$$

Model (3) identifies A and B as alternative direct causes of C and indirect causes of E, moreover C and D are claimed to be alternative direct causes of E. All of these claims are also made by (2). Analogously, model (4) expresses all that is stated by (2), and adds G to the path of D leading to E. The causal claims entailed by (3) thus constitute a subset of the claims entailed by (2), which in turn are contained in the claims of (4).[3] All three models describe properties of one and the same underlying causal structure at different degrees of detail and relative to different sets of data $\delta_{(2)}$, $\delta_{(3)}$ and $\delta_{(4)}$.

2.3 Data, consistency, coverage

CCMs analyze *configurational data* δ that have the form of $m \times k$ matrices, where m is the number of units of observation (cases) and k is the number of factors in δ. We subsequently refer to the set of factors $\mathbf{F}$ in an analyzed δ as the *factor frame* of the analysis. While QCA requires that $\mathbf{F}$ be partitioned—prior to the analysis—into a first subset $\{Y\}$ comprising exactly one endogenous factor and a second subset $\mathbf{F} \setminus \{Y\}$ comprising all exogenous factors of the analysis, CNA can dispense with such a partition. If prior causal knowledge is available as to what factors in $\mathbf{F}$ are possible effects and what factors can be excluded as effects, this information can be given to CNA via an optional argument called a *causal ordering*. A causal ordering is a relation $X_i \prec X_j$ defined on the elements of $\mathbf{F}$ entailing that X_j cannot be a cause of X_i (e.g. because X_i is instantiated temporally before X_j). If an ordering is provided, CNA only searches for Boolean models in accordance with the ordering; if no ordering is provided, CNA treats all values of the factors in $\mathbf{F}$ as potential outcomes and explores whether a causal model for them can be inferred from δ.

As real-life data tend to feature noise induced by uncontrolled (unmeasured) causes of endogenous factors, it often happens that no configuration of factor values is strictly sufficient or necessary for a given outcome Y. To still extract some causal information from such data, Ragin (2006) has introduced so-called *consistency* and *coverage* measures (with values from the interval $[0, 1]$) into the QCA protocol. Both of these measures are also ser-

[3]The subset relations among the *causal claims* entailed by Boolean models of varying complexity must not be confused with the subset relations among *the sets of cases* defined by conjunctions and disjunctions of varying complexity. (3) only ascribes causal relevance to A for C, whereas (4) ascribes causal relevance to both A and b for C and, hence, entails more causal relevancies than (3). By contrast, the set of cases exhibiting the configuration $A{*}b$ is a subset of the set of cases exhibiting A.

viceable to the purposes of CNA. Informally put, *consistency* reproduces the degree to which the behavior of an outcome obeys a corresponding sufficiency or necessity relationship or a whole model, whereas *coverage* reproduces the degree to which a sufficiency or necessity relationship or a whole model accounts for the behavior of the corresponding outcome. As the implication operator underlying the notions of sufficiency and necessity is defined differently in classical and in fuzzy logic, the two measures are defined differently for crisp-set (cs) and multi-value data, on the one hand, and fuzzy-set (fs) data, on the other. *Cs-consistency* (con^{cs}) of $X \rightarrow Y$ is defined as the number of cases in δ featuring $X{*}Y$ divided by the number of cases featuring X, and *cs-coverage* (cov^{cs}) of $X \rightarrow Y$ amounts to the number of cases in δ featuring $X{*}Y$ divided by the number of cases featuring Y. By contrast, *fs-consistency* (con^{fs}) and *fs-coverage* (cov^{fs}) of $X \rightarrow Y$ are defined as follows, where n is the number of cases in δ:

$$con^{fs}(X \rightarrow Y) = \frac{\sum_{i=1}^{n} \min(X_i, Y_i)}{\sum_{i=1}^{n} X_i} \qquad cov^{fs}(X \rightarrow Y) = \frac{\sum_{i=1}^{n} \min(X_i, Y_i)}{\sum_{i=1}^{n} Y_i}$$

Although defined differently, the cs and fs variants of these measures are not logically independent, as con^{cs} and cov^{cs} are special cases of con^{fs} and cov^{fs} where all membership scores are equal to 0 or 1. Therefore, whenever the context renders it sufficiently clear what the relevant data structure is, we will henceforth not explicitly distinguish between the cs and fs measures but simply speak of consistency (con) and coverage (cov).

If no strictly Boolean relations of sufficiency and necessity and, thus, no minimal theories with $con = 1$ and $cov = 1$ can be inferred from δ, CNA invites its users to lower the con and cov thresholds con_t and cov_t. For example, by lowering con_t^{cs} to 0.8, CNA is given permission to treat X as sufficient for Y, even though in 20% of the cases X is not associated with Y. Or by lowering cov_t^{fs} to 0.8, CNA is allowed to treat X as necessary for Y, even though the sum of the membership scores in Y over all cases in δ exceeds the sum of the membership scores in $\min(X, Y)$ by 20%.

Lowering con_t and cov_t must be done with great caution, for the lower these thresholds, the higher the chance that causal fallacies are committed. However, in QCA, it is common to only impose lowest bounds for the consistency of configurations of all exogenous factors, so-called *minterms*. For instance, Schneider and Wagemann (2010) recommend a lowest bound of $con_t = 0.75$ for minterm sufficiency. Yet, this approach does not guarantee that the consistencies of the issued minimally sufficient conditions (or *prime implicants*, as they are called in QCA) and of the resulting Boolean models are also above the chosen threshold. Accordingly, it often happens that the models output by QCA do not meet the consistency

threshold set by the user (cf. the replication script for examples). Moreover, it is common QCA practice not to require lowest bounds for the coverage of resulting models. In consequence, QCA models frequently cover less than half of the occurrences of the outcome in δ.

In CNA, the consistency and coverage standards are higher—for two reasons. First, the sufficient conditions that are ultimately causally interpreted by CCMs are not minterms (which are mere intermediate calculation devices for QCA) but redundancy-free conditions contained in Boolean models. Hence, consistency thresholds must be imposed on the latter, not on the former. Second, a model's coverage being low means that it only accounts for few instances of an outcome in δ. Or differently, in many cases in δ where the outcome is present there are causes at work that are not contained in the factor frame $\mathbf{F}$. However, unmeasured causes tend to confound δ—in particular, when they are associated with both exogenous and endogenous factors in $\mathbf{F}$. The possible existence of confounders casts doubts on the causal interpretability of all dependencies manifest in δ, for uncontrolled causes might be covertly responsible for them. That is, the more likely it is that the data are confounded, the less reliable a causal interpretation of resulting models becomes. In CNA, therefore, the coverage scores of the models are used as a measure for the likelihood of confounding. The higher the coverage, the less likely it is that we are facing data confounding, the more reliable a causal interpretation of issued models becomes. Hence, CNA aims to maximize both the consistency and coverage scores; neither of the corresponding thresholds should be lowered below 0.75.[4]

2.4 Homogeneity, model ambiguities, correctness

While high consistency and coverage scores increase the reliability of resulting models, they do not guarantee their correctness. To get a clear understanding of the scope, inferential potential, and limitations of CNA, this subsection spells out what can and what cannot be expected of CNA. More specifically, we explain what it means for the output of CNA to be correct and under what conditions CNA will certainly produce a correct output. Very generally put, to say that CNA—or any other method for that matter—is a *correct* procedure of causal inference means that the causal conclusions it draws from data δ are true of the δ-generating causal structure Δ.

[4]Note that requiring high model coverage does not amount to requiring that a study explains *all* occurrences of an outcome, as Thiem (2015, 728) contends; rather, it simply means that a study should explain all of an outcome's occurrences *in the data*.

This general characterization of correctness calls for two specifications. First, no method can be expected to systematically infer true causal models from deficient data. Whether data meet required quality standards depends on whether they faithfully reflect the causal structure that generated them. But since this structure is unknown in real-life discovery contexts, data quality cannot be assessed analytically, rather, it must be imposed by assumption (Cartwright 1989, 55-90). Even heuristics designed to ensure compliance with these assumptions, such as randomization and experimental control, cannot eliminate the risk of insufficient data quality. Accordingly, all procedures of causal inference come with a set of background assumptions, and can only be expected to produce correct results provided these assumptions are satisfied.[5]

While in most methodological traditions, the details of these background assumptions are thoroughly investigated and debated, the CCM literature has unfortunately largely sidestepped this important issue so far. We cannot exhaustively fill this gap here (which would require a study in its own right), but still want to provide one background assumption—the configurational *homogeneity* assumption—which is sufficient to ensure the correctness of CCMs by ensuring that the analyzed data are not confounded (cf. Baumgartner 2009a).[6] Generally put, configurational data are *confounded* iff unmeasured causes change between observed cases in such a way that resulting differences in the outcomes appear to be due to the measured factors, whereas they are actually due to the changing unmeasured causes. Factors that can induce confounding are unmeasured causes of a scrutinized outcome Y that change the value of Y in a way that is not mediated through the measured factors in an analyzed frame $\mathbf{F}$, i.e. causes of Y that are connected to Y on at least one causal path that does not go through the elements of $\mathbf{F}$ and to which we shall, hence, refer as *off*-$\mathbf{F}$-*path* causes of Y.[7] Changes in off-$\mathbf{F}$-path causes of Y can bring about changes in Y that are erroneously ascribed to a measured factor that merely happens to co-vary in its values with Y without being causally relevant to Y. Configurational data δ are not confounded if all off-$\mathbf{F}$-path causes of Y remain constant across all cases in δ. Accordingly, an assumption

[5]For instance, regression analytic methods impose the *Gauss-Markov* assumptions (Gelman and Hill 2007, 45-47), and Bayesian network methods rely on the *Causal Markov* and *Faithfulness* assumptions (Spirtes et al. 2000, 29-31).

[6]We have to leave it to future research to determine whether the homogeneity assumption is also necessary for that purpose, or whether there exist alternative, possibly weaker assumptions that could likewise guarantee CNA's correctness. Moreover, note that data confounding is, of course, not the only data deficiency that can induce causal fallacies, errors of data collection (e.g. measurement error or selection bias) being another common type of data deficiency. For the purposes of this paper, we bracket errors of data collection by assuming that data have been faultlessly collected.

[7]This terminology is derived from Woodward's (2003, 59-60) notion of an *off-path variable*.

that is sufficient to exclude confounding stipulates that δ are *homogenous* in the following sense:

Configurational Homogeneity (CH): Configurational data δ for an outcome Y over a factor frame $\mathbf{F}$ are homogenous iff every off-$\mathbf{F}$-path cause of Y remains constant in all cases in δ.

Requiring δ to be homogenous in the sense of **CH** amounts to a strong assumption that may be difficult to justify in observational studies. However, on a par with background assumptions in other methodological frameworks, the function of **CH** merely is to *guarantee* the correctness of a CCM as CNA.[8] If data δ are homogenous, it follows that all observed differences in the outcomes must be due to variations of the measured factors, which, in turn, ensures that CNA cannot commit fallacies by ascribing the difference-making relations it uncovers in δ to causal influences of the measured factors. At the same time, it is important to emphasize that, if CNA is applied to data that violate **CH**, it does not automatically follow that incorrect models will be generated, rather, it only follows that correctness is no longer guaranteed. In fact, section 4 will show that CNA frequently produces correct models even in data scenarios violating **CH**.

The second necessary specification of the rough characterization of the correctness criterion concerns the phenomenon of model ambiguities. There often exist multiple causal models that fit analyzed data equally well, to the effect that the data underdetermine their own causal modeling. Model ambiguities are a very common phenomenon in all methodological traditions (Simon 1954; Spirtes et al. 2000, 59-72; Baumgartner and Thiem 2015).[9] Of course, CNA—on a par with any other method—cannot disambiguate what is empirically underdetermined. Rather, it must draw those and only those causal conclusions for which the analyzed data *de facto* contain evidence. In cases of empirical underdetermination it must, therefore, render transparent all data-fitting models (and leave the disambiguation up to the analyst). Not all of these models may correctly reflect the data-generating causal structure, but in order for CNA to pass as a correct method of causal inference the data-generating

[8] Models output by QCA are likewise only guaranteed to be correct if the data comply with **CH** (for an extended argument to that effect cf. Baumgartner and Thiem 2017). Unfortunately, many representatives of QCA are not sufficiently transparent with respect to the assumptive basis of their method.

[9] As shown by Baumgartner and Thiem (2015), model ambiguities are much more frequent in configurational causal modeling than is typically acknowledged. In particular, applications of QCA are affected by a widespread practice of model-underreporting, one main reason being that the dominant QCA computer programs—as fs/QCA (Ragin and Davey 2016), Tosmana (Cronqvist 2017), and QCAGUI (Duşa 2007)—regularly fail to uncover the whole model space, even for ideal data. The only currently available QCA program not affected by this problem is QCApro (Thiem 2016).

structure must be truthfully reflected by at least one generated model (Spirtes et al. 2000, 81). If CNA outputs, say, three models $\mathbf{m}_i$, $\mathbf{m}_j$, and $\mathbf{m}_k$, it determines that the data-generating structure has the form of $\mathbf{m}_i$ *or* that of $\mathbf{m}_j$ *or* that of $\mathbf{m}_k$. Such a disjunction is true iff at least one disjunct is true.

Overall, for CNA—or any other CCM—to be a correct method of causal inference it is required that at least one model inferred from homogenous data truthfully reflects the Boolean causal properties of the data-generating structure. More explicitly:

Configurational Correctness (CC): A configurational comparative method $\mathcal{P}$ is a correct procedure of causal inference iff, whenever $\mathcal{P}$ infers a set of models $\mathbf{M}$ from data δ which comply with **CH**, (at least) one model $\mathbf{m}_i \in \mathbf{M}$ satisfies the following four conditions:

(1) all values of exogenous factors contained in $\mathbf{m}_i$ are causally relevant for the corresponding outcome in the δ-generating structure Δ;

(2) if two values of X and Y are contained in two different disjuncts in $\mathbf{m}_i$, then X and Y are located on two different causal paths in Δ;

(3) if two values of X and Y are contained in the same conjunct in $\mathbf{m}_i$, then X and Y are part of the same complex cause in Δ;

(4) if two values of X and Y are two links of a causal chain in $\mathbf{m}_i$, then X and Y are located on a causal chain in Δ.

We contend that CNA is correct in the sense defined by **CC** and will provide substantive evidence for this claim in section 4. For now, two aspects of the claim that CNA is a correct procedure deserve separate emphasis. First, that CNA is correct does not entail that it infers a set of causal models from every data input. Empirical data may be insufficient to warrant any causal inference or to meet predefined thresholds of model fit. Whenever CNA abstains from such an inference, it cannot commit a causal fallacy. By extension, correctness cannot be violated. Users of CCMs in general and of CNA in particular should keep in mind that configurational causal modeling imposes very high quality standards on the processed data. If these standards are not met, a reliable CCM must abstain from drawing causal inferences. Accordingly, when applied to real-life data featuring various sorts of deficiencies, CNA will often output no model at all. If that happens, the analyst should go back to the data and attempt to improve its quality (e.g. by controlling for unmeasured causes) instead of lowering the standards imposed on data quality or the consistency and coverage thresholds. It is better not to draw a causal inference than to draw a hazardous one. As already anticipated in

the previous section, CNA adopts a much more risk-averse approach in dealing with data deficiencies than QCA. While the latter does not impose a coverage threshold at all and often causally interprets minimally sufficient conditions that do not meet the consistency threshold, the former uses both consistency and coverage as *authoritative* model building criteria such that, if they are not met, CNA abstains from a causal inference.

Second, that CNA is a correct method does not entail that it always *completely* uncovers the properties of the data-generating structure Δ. Real-life data tend to be fragmentary, meaning they do not contain all empirically possible configurations of the factors in an analyzed frame, that is, all configurations that are compatible with Δ.[10] Fragmentary data may not contain evidence for certain features of Δ, and no method can compensate for lacking evidence in the data. Correctness merely demands that, if CNA outputs a set $\mathbf{M}$, then at least one model $\mathbf{m}_i \in \mathbf{M}$ be such that all causal properties represented by $\mathbf{m}_i$ truthfully reflect *some* causal properties of Δ. At the same time, if CNA is given exhaustive data featuring *all* empirically possible configurations, CNA should completely uncover Δ. That is, *completeness* is imposed as a conditional criterion: if CNA is given exhaustive data in compliance with **CH**, the Boolean causal properties represented by at least one model $\mathbf{m}_i \in \mathbf{M}$ truthfully reflect *all* Boolean causal properties of Δ.[11]

Since data fragmentation is ubiquitous in observational studies, procedures employed in this domain usually will only uncover a proper part of Δ. Still, if δ is fragmented, CNA will uncover all those parts of Δ for which δ contains evidence, no fewer and no more. More specifically, although CNA is not unconditionally complete, it is unconditionally *informative* in the following sense: all and only those Boolean causal properties of Δ for which δ contains evidence are truthfully reflected by at least one model $\mathbf{m}_i \in \mathbf{M}$.

[10] *Data fragmentation*, as we use the term here, is related but not synonymous to *limited diversity*, a concept which is known from the QCA literature (e.g. Ragin 2008, 147-148). QCA assumes that configurational data feature exactly one endogenous factor and that all exogenous factors are causally independent of one another. QCA-processed data is said to be limitedly diverse iff they do not contain all *logically possible* configurations of the exogenous factors. CNA, by contrast, allows for the factors that are exogenous with respect to some ultimate outcome to be mutually causally dependent, in which case not all logically possible configurations are also empirically possible. Accordingly, we say that data are fragmented iff they do not contain all *empirically possible* configurations.

[11] In Baumgartner (2009a), an assumption of *empirical exhaustiveness* (PEX) is introduced to ensure that CNA-processed data is non-fragmentary and that Δ could be completely uncovered. We dispense with that assumption here. As a result, CNA will not always completely uncover Δ.

3 Generalizing the CNA algorithm

3.1 Top-down vs. bottom-up search

The goal of CCMs is to infer Boolean causal models from configurational data. The previous section has shown that Boolean functions are amenable to a causal interpretation only if they identify redundancy-free sufficient and necessary conditions, and thus amount to minimal theories that reach imposed consistency (con_t) and coverage (cov_t) thresholds.

There exist (at least) two different strategies for building minimal theories: they can be built from the *top down* or from the *bottom up*. The top-down approach proceeds as follows. First, complete sufficient minterms are identified that meet con_t; second, elements are eliminated as redundant as long as the remaining conditions continue to satisfy con_t; third, the minimally sufficient conditions are disjunctively combined to necessary conditions that meet cov_t; fourth, elements are eliminated as redundant that are not required to satisfy cov_t. By contrast, the bottom-up approach starts with single factor values and tests whether they meet con_t; if that is not the case, it proceeds to test conjunctions of two factor values, then to conjunctions of three, and so on. Whenever a conjunction meets con_t (and no proper part of it has previously been identified to meet con_t), it is automatically redundancy-free, that is, a minimally sufficient condition (MSC), meaning that supersets of it do not need to be tested for sufficiency any more. Then, the bottom-up approach tests whether single MSC meet cov_t; if not, it proceeds to disjunctions of two, then to disjunctions of three, and so on. Whenever a disjunction meets cov_t (and no proper part of it has previously been identified to meet cov_t), it is automatically redundancy-free, viz. a minimally necessary condition, and thus supersets of it do not need to be tested for necessity.

Both QCA and the original variant of csCNA adopt versions of the top-down approach—albeit they algorithmically implement this approach very differently (cf. Baumgartner 2014). By contrast, the generalization of CNA developed here reverses the direction of model building. Prima facie, it might be thought that it does not matter much whether models are built from the top down or from the bottom up because both directions should ultimately lead to the same results. Although that is indeed the case for some data types, in particular for ideal data, it does not hold generally. For instance, when applied to data that do not allow for modeling outcomes with perfect consistency, it can happen that—contrary to the bottom-up approach—the top-down approach does not succeed in eliminating all redundancies from sufficient conditions. The reason is that when building models from the top down it is (implicitly) presumed that consistency threshold violations are *monotonic* in the fol-

A	B	C	D	n
1	1	1	1	3
1	1	1	0	1
1	1	0	1	5
1	1	0	0	2
1	0	1	0	1
1	0	0	1	7
0	1	1	0	4
1	0	0	0	1
0	0	0	0	3

Table 1: Condensed representation of a configurational data set, where the right-most column lists the number of cases featuring a configuration and D is the outcome.

lowing sense: if a single factor C cannot be eliminated from a sufficient condition $A{*}B{*}C$ because $A{*}B$ alone does not meet con_t, then C *plus* some further factor from $A{*}B$ cannot be eliminated either. In consequence, if eliminating C from $A{*}B{*}C$ leads to a violation of con_t, the top-down approach concludes that C is needed to account for the outcome, meaning C is a difference-maker of the outcome. That conclusion, however, is not valid, because consistency threshold violations are not monotonic in the above sense.

To see this, consider the data matrix in Table 1, in which the following consistencies hold:

$$con(A{*}B{*}C \rightarrow D) = {}^{3}\!/_{4} = 0.75$$
$$con(A{*}B \rightarrow D) = {}^{8}\!/_{11} = 0.73$$
$$con(A \rightarrow D) = {}^{15}\!/_{20} = 0.75$$

That is, if con_t is set to 0.75, the condition $A{*}B{*}C$, which is sufficient for D with consistency 0.75, satisfies the threshold. By contrast, $A{*}B$, which results from $A{*}B{*}C$ by eliminating C, falls short of con_t. Nonetheless, further eliminating B lifts the remaining condition above con_t again, as A alone is sufficient for D with consistency 0.75. That means, while C initially appears to be a non-redundant element of the sufficient condition $A{*}B{*}C$, it turns out to be redundant after all. The top-down approach, however, only tests the removability of single factors at a time and infers that a condition is redundancy-free if removing single factors would push that condition below con_t. Therefore, at a consistency threshold of 0.75, a procedure that adopts the top-down approach as QCA issues model (5) for Table 1.

$$A{*}b{*}c \ + \ A{*}B{*}C \rightarrow D \qquad con = 0.83;\ cov = 0.67 \tag{5}$$

In contrast, by first testing whether single factors meet con_t, the bottom-up approach directly finds that A itself is sufficient for D. Moreover, it turns out that A is also necessary for D in Table 1, as it accounts for D with perfect coverage. Overall, at a consistency threshold of 0.75, a procedure that builds models from the bottom up issues model (6).

$$A \leftrightarrow D \qquad con = 0.75;\ cov = 1 \tag{6}$$

Model (6) is preferable over model (5), for two reasons. First, the product of consistency and coverage, which is commonly used as a measure for overall model fit, is significantly higher for (6) than for (5). Second, model (6) only ascribes causal relevance to A, whereas (5) also determines B, C and their negations to be causes of D, even though the data in Table 1 do not contain evidence that B, C or their negations actually would make a difference to D. In sum, when applied to noisy data the top-down approach runs a risk of drawing causal inferences that go beyond the evidence contained in the data, because it may not succeed in completely eliminating redundant elements from causal models. To avoid this problem, the generalization of CNA developed in this paper builds models from the bottom up.

3.2 The essentials of the CNA algorithm

The generalized CNA algorithm takes as mandatory inputs (i) a data set δ, (ii) con_t and cov_t thresholds, and (iii) an upper bound called $maxstep$ for the maximal complexity of atomic solution formulas (atomic causal models) to be built. $Maxstep$ serves the pragmatic purpose of keeping the search space computationally tractable in reasonable time. The user can set it to any complexity level if computational time is not an issue. Optionally, CNA can be given a causal ordering as introduced in section 2.3. Moreover, an additional consistency threshold called $con.msc_t$ can be defined that is only imposed on the consistency to be reached by minimally sufficient conditions. The default setting is $con.msc_t = con_t$, in which case minimally sufficient conditions, atomic and complex causal models all have to meet the same thresholds. If $con.msc_t \neq con_t$, minimally sufficient conditions must comply with $con.msc_t$, while causal models are required to satisfy con_t. As illustrated in the replication script, setting $con.msc_t \neq con_t$ can enhance the informativeness of CNA's output in certain cases. In the ensuing discussion, however, we will standardly assume $con.msc_t = con_t$.

	A	B	C	D
c_1	0	0	0	0
c_2	0	1	0	0
c_3	1	1	0	0
c_4	0	0	1	0
c_5	1	0	0	1
c_6	1	0	1	1
c_7	0	1	1	1
c_8	1	1	1	1

(a) cs data

	A	B	C	D
c_1	1	3	3	1
c_2	2	2	1	2
c_3	2	1	2	2
c_4	2	2	2	2
c_5	3	3	3	2
c_6	2	4	3	2
c_7	1	3	3	3
c_8	1	4	3	3

(b) mv data

	A	B	C	D	E
c_1	0.17	0.02	0.15	0.26	0.09
c_2	0.97	0.23	0.73	0.08	0.10
c_3	0.10	0.72	0.61	0.38	0.08
c_4	0.64	0.73	0.82	0.12	0.66
c_5	0.11	0.30	0.06	0.99	0.78
c_6	0.69	0.23	0.91	0.98	0.84
c_7	0.31	0.80	0.62	0.65	0.74
c_8	0.65	0.87	0.92	0.82	0.85

(c) fs data

Table 2: Data types that can be analyzed by CNA.

Contrary to QCA, which first transforms the data into an intermediate calculative device called a *truth table*, the CNA algorithm operates directly on the data.[12] Data processed by CNA can either be of type "crisp-set" (cs), "multi-value" (mv) or "fuzzy-set" (fs). Examples of each data type are given in Table 2. In what follows, we first discuss the generalized CNA algorithm in the abstract, using the explicit 'Variable=value' notation, and then we illustrate its procedural steps on the basis of the fuzzy-set data in Table 2(c).

CNA causally models configurational data δ over a factor frame $\mathbf{F}$ in four stages:

Stage 1 CNA uses a provided ordering to build a set of potential outcomes $\mathbf{O} = \{O_1 = \omega_1, \ldots, O_n = \omega_n\}$ from the factor frame $\mathbf{F}$ and to assign a set of potential cause factors $\mathbf{C}_{O_i}$ in $\mathbf{F} \setminus O_i$ to every element $O_i = \omega_k$ of $\mathbf{O}$. If no ordering is provided, all value assignments to all elements of $\mathbf{F}$ are treated as possible outcomes in case of mv-data, whereas in case of cs- and fs-data CNA sets $\mathbf{O}$ equal to $\{O_1 = 1, \ldots, O_n = 1\}$, where O_1 to O_n are all the factors in $\mathbf{F}$.

Stage 2 CNA attempts to build a set $\mathbf{MSC}_{O_i=\omega_k}$ of minimally sufficient conditions that meet con_t for each $O_i = \omega_k \in \mathbf{O}$. To this end, it first checks for each value assignment $X_h = \chi_j$ of each element of $\mathbf{C}_{O_i}$, such that $X_h = \chi_j$ has a membership score above 0.5 in at least one case in δ, whether the consistency of $X_h = \chi_j \;\rightarrow\; O_i = \omega_k$ in δ meets con_t, i.e. whether $con(X_h = \chi_j \;\rightarrow\; O_i = \omega_k) \geq con_t$. If, and only if, that is the case, CNA puts $X_h = \chi_j$ into the set $\mathbf{MSC}_{O_i=\omega_k}$. Next, CNA checks for each conjunction of two factor values $X_m = \chi_j * X_n = \chi_l$ from $\mathbf{C}_{O_i}$, such that $X_m = \chi_j * X_n = \chi_l$ has a membership score above 0.5 in at least one case in δ and no part of $X_m = \chi_j * X_n = \chi_l$

[12]The cna R-package provides a function called *truthTab* which has a mere representational purpose: it transparently rearranges the data for the user to review. The object produced by *truthTab* must not be confused with a QCA truth table.

is already contained in $\mathbf{MSC}_{O_i=\omega_k}$, whether $con(X_m = \chi_j * X_n = \chi_l \rightarrow O_i = \omega_k) \geq con_t$. If, and only if, that is the case, CNA puts $X_m = \chi_j * X_n = \chi_l$ into the set $\mathbf{MSC}_{O_i=\omega_k}$. Next, conjunctions of three factor values with no parts already contained in $\mathbf{MSC}_{O_i=\omega_k}$ are tested, then conjunctions of four factor values, etc., until either all logically possible conjunctions of the elements of $\mathbf{C}_{O_i}$ have been tested or $maxstep$ is reached. Every non-empty $\mathbf{MSC}_{O_i=\omega_k}$ is passed on to the third stage.

Stage 3 CNA attempts to build a set $\mathbf{ASF}_{O_i=\omega_k}$ of atomic solution formulas (atomic causal models) for every $O_i = \omega_k \in \mathbf{O}$, which has a non-empty $\mathbf{MSC}_{O_i=\omega_k}$, by disjunctively concatenating the elements of $\mathbf{MSC}_{O_i=\omega_k}$ to minimally necessary conditions of $O_i = \omega_k$ that meet cov_t. To this end, it first checks for each single condition $\Phi_h \in \mathbf{MSC}_{O_i=\omega_k}$ whether the coverage of $\Phi_h \rightarrow O_i = \omega_k$ in δ meets cov_t, i.e. $cov(\Phi_h \rightarrow O_i = \omega_k) \geq cov_t$. If, and only if, that is the case, CNA puts Φ_h into the set $\mathbf{ASF}_{O_i=\omega_k}$. Next, CNA checks for each disjunction of two conditions $\Phi_m + \Phi_n$ from $\mathbf{MSC}_{O_i=\omega_k}$, such that no part of $\Phi_m + \Phi_n$ is already contained in $\mathbf{ASF}_{O_i=\omega_k}$, whether $cov(\Phi_m + \Phi_n \rightarrow O_i = \omega_k) \geq cov_t$. If, and only if, that is the case, CNA puts $\Phi_m + \Phi_n$ into the set $\mathbf{ASF}_{O_i=\omega_k}$. Next, disjunctions of three conditions from $\mathbf{MSC}_{O_i=\omega_k}$ with no parts already contained in $\mathbf{ASF}_{O_i=\omega_k}$ are tested, then disjunctions of four conditions, etc., until either all logically possible disjunctions of the elements of $\mathbf{MSC}_{O_i=\omega_k}$ have been tested or $maxstep$ is reached. Every non-empty $\mathbf{ASF}_{O_i=\omega_k}$ is passed on to the fourth stage.

Stage 4 CNA attempts to build a set $\mathbf{CSF}_{\mathbf{O}}$ of complex solution formulas (complex causal models) encompassing all elements of $\mathbf{O}$. To this end, CNA constructs all logically possible conjunctions of exactly one element from every non-empty $\mathbf{ASF}_{O_i=\omega_k}$. If there is only one non-empty set $\mathbf{ASF}_{O_i=\omega_k}$, that is, if only one potential outcome can be modeled as an actual outcome, the set of complex solution formulas $\mathbf{CSF}_{\mathbf{O}}$ is identical to $\mathbf{ASF}_{O_i=\omega_k}$.

To illustrate all four stages, let us now apply CNA to Table 2(c). We set $con_t = 0.8$ and $cov_t = 0.9$ and execute the algorithm in the most general manner by not providing an ordering. 2(c) contains data of type fs, meaning that the values in the data matrix are interpreted as membership scores in fuzzy sets. As is customary for this data type, we use uppercase letters for membership in a set and lowercase letters for non-membership. In the absence of an ordering, the first stage determines the set of potential outcomes to be $\mathbf{O} = \{A, B, C, D, E\}$, that is, the presence of each factor in 2(c) is treated as a potential

outcome. Moreover, all other factors are potential cause factors of every element of **O**, hence, $\mathbf{C}_A = \{B, C, D, E\}$, $\mathbf{C}_B = \{A, C, D, E\}$, $\mathbf{C}_C = \{A, B, D, E\}$ etc.

As the construal of minimally sufficient conditions in stage 2 and of atomic solution formulas in stage 3 requires testing a multitude of conditions for con_t and cov_t compliance, these stages cannot be exhaustively illustrated here (for more details cf. the replication script). Figure 1 contains a transcript of how CNA arrives at $\mathbf{MSC}_C$, $\mathbf{MSC}_D$ and $\mathbf{MSC}_E$ in stage 2 and at $\mathbf{ASF}_C$ and $\mathbf{ASF}_E$ in stage 3. Overall, stage 2 succeeds in building non-empty sets of minimally sufficient conditions for all elements of **O**: $\mathbf{MSC}_A = \{b{*}C, d{*}E\}$, $\mathbf{MSC}_B = \{a{*}C, A{*}E, d{*}E\}$, $\mathbf{MSC}_C = \{A, B, d{*}E\}$, $\mathbf{MSC}_D = \{E, a{*}C\}$, $\mathbf{MSC}_E = \{D, A{*}B\}$. But only the elements of $\mathbf{MSC}_C$ and $\mathbf{MSC}_E$ can be disjunctively combined to atomic solution formulas that meet cov_t in stage 3: $\mathbf{ASF}_C = \{A + B \leftrightarrow C\}$ and $\mathbf{ASF}_E = \{D + A{*}B \leftrightarrow E\}$. For the other three factors in **O** the coverage threshold of 0.9 cannot be satisfied. CNA therefore abstains from issuing causal models for A, B and D.

Finally, stage 4 conjunctively concatenates $\mathbf{ASF}_C$ and $\mathbf{ASF}_E$ to the following complex solution formula $\mathbf{CSF}_\mathbf{O}$, which constitutes CNA's final causal model for Table 2(c):

$$(A \ + \ B \leftrightarrow C) * (D \ + \ A{*}B \leftrightarrow E) \qquad con = 0.808;\ cov = 0.925 \tag{7}$$

Two features of this generalized version of the CNA algorithm deserve (re-)emphasis. First, while the computational cores of configurational methods that build models from the top down are constituted by procedures for redundancy elimination that turn maximal into minimal sufficient and necessary conditions, all conditions that are found to comply with con_t and cov_t by CNA are automatically redundancy-free. That is, CNA directly identifies *minimally* sufficient and necessary conditions in the data, rendering redundancy elimination itself redundant. Second, whereas QCA has to dichotomize fs data in a truth table before processing it, CNA processes fs data in the very same vein as cs and mv data, viz. by building all viable conjunctions and disjunctions of potential causes and systematically testing for con_t and cov_t compliance. By directly applying the same algorithm to all configurational data types, CNA renders the detour via truth tables redundant.

4 Evaluation and comparison

Before a new method can be applied in real-life studies, it must, on the one hand, be shown that the method correctly analyzes data that are ideal by its own standards, and, on the other, an estimate should be provided of how the method performs under different constellations of

```
OUTCOME: C
--------------------------
msc, step 1
      con     cov min.suff
A   0.915   0.691      YES
a   0.571   0.517        .
B   0.864   0.699      YES
b   0.632   0.537        .
D   0.734   0.651        .
d   0.772   0.595        .
E   0.797   0.685        .
e   0.637   0.510        .
--------------------------
msc, step 2
       con     cov min.suff
a*b   0.520   0.297        .
a*d   0.740   0.355        .
a*D   0.676   0.419        .
a*e   0.609   0.365        .
a*E   0.706   0.394        .
b*d   0.740   0.371        .
b*D   0.705   0.371        .
b*e   0.673   0.411        .
b*E   0.726   0.353        .
d*e   0.747   0.471        .
d*E   1.000   0.309      YES
D*E   0.783   0.562        .
-----------------------------
msc, step 3
         con    cov min.suff
a*b*d   0.649  0.226        .
a*b*D   0.631  0.266        .
a*b*e   0.604  0.266        .
a*b*E   0.646  0.243        .
a*d*e   0.724  0.326        .
a*D*E   0.689  0.344        .
b*d*e   0.740  0.371        .
b*D*E   0.705  0.317        .
-----------------------------
msc, step 4
           con   cov min.suff
a*b*d*e   0.649 0.226       .
a*b*D*E   0.614 0.212       .
=============================
asf, initial
             con    cov
A + B + d*E 0.850  0.925
--------------------------
asf, step 1
        con    cov  min.nec
A     0.915  0.691        .
B     0.864  0.699        .
d*E   1.000  0.309        .
-------------------------------
asf, step 2
           con    cov min.nec
A + B     0.850  0.925     YES
A + d*E   0.916  0.703       .
B + d*E   0.866  0.714       .
-------------------------------
asf, end
```

```
OUTCOME: D
--------------------------
msc, step 1
      con     cov min.suff
A   0.612   0.521        .
a   0.686   0.699        .
B   0.667   0.607        .
b   0.620   0.593        .
C   0.651   0.734        .
c   0.732   0.544        .
E   0.836   0.808      YES
e   0.422   0.381        .
-----------------------------
msc, step 2
       con     cov min.suff
a*b   0.738   0.474        .
A*b   0.656   0.374        .
a*B   0.770   0.486        .
A*B   0.707   0.379        .
a*c   0.777   0.521        .
A*C   0.649   0.505        .
a*C   0.811   0.472      YES
a*e   0.547   0.369        .
A*e   0.525   0.269        .
b*c   0.683   0.423        .
b*C   0.691   0.418        .
B*C   0.691   0.544        .
b*e   0.493   0.339        .
B*e   0.662   0.325        .
c*e   0.637   0.348        .
C*e   0.553   0.318        .
-----------------------------
msc, step 3
         con    cov min.suff
a*b*c   0.736  0.411        .
A*b*C   0.657  0.357        .
A*B*C   0.701  0.367        .
a*b*e   0.660  0.327        .
A*b*e   0.560  0.250        .
a*B*e   0.705  0.313        .
a*c*e   0.692  0.336        .
A*C*e   0.554  0.252        .
b*c*e   0.613  0.311        .
b*C*e   0.596  0.276        .
B*C*e   0.672  0.287        .
-------------------------------
msc, step 4
           con   cov min.suff
a*b*c*e   0.670 0.299       .
A*b*C*e   0.556 0.233       .
=============================
asf, intial
         con    cov
E + a*C 0.808  0.893
No asf's!
```

```
OUTCOME: E
--------------------------
msc, step 1
      con     cov min.suff
A   0.734   0.645        .
a   0.617   0.650        .
B   0.764   0.720        .
b   0.571   0.565        .
C   0.685   0.797        .
c   0.560   0.430        .
D   0.809   0.836      YES
d   0.401   0.360        .
-----------------------------
msc, step 2
       con     cov min.suff
a*b   0.659   0.437        .
A*b   0.684   0.403        .
a*B   0.763   0.498        .
A*B   0.934   0.517      YES
a*c   0.596   0.413        .
A*C   0.787   0.632        .
a*C   0.763   0.459        .
a*d   0.485   0.271        .
A*d   0.609   0.345        .
b*c   0.574   0.367        .
b*C   0.656   0.417        .
B*C   0.777   0.633        .
b*d   0.372   0.217        .
B*d   0.657   0.343        .
c*d   0.446   0.220        .
C*d   0.519   0.360        .
-----------------------------
msc, step 3
         con    cov min.suff
a*b*c   0.607  0.350        .
A*b*C   0.695  0.391        .
a*B*C   0.766  0.423        .
a*b*d   0.494  0.200        .
A*b*d   0.539  0.217        .
a*B*d   0.660  0.254        .
a*c*d   0.467  0.203        .
A*C*d   0.668  0.345        .
a*C*d   0.655  0.271        .
b*c*d   0.427  0.184        .
b*C*d   0.503  0.217        .
B*C*d   0.660  0.343        .
-------------------------------
msc, step 4
           con   cov min.suff
a*b*c*d   0.448 0.167       .
A*b*C*d   0.559 0.217       .
a*B*C*d   0.665 0.254       .
=============================
asf, intial
         con    cov
D + A*B 0.808  0.966
-------------------------------
asf, step 1
        con    cov min.nec
D     0.808  0.837       .
A*B   0.934  0.517       .
-------------------------------
asf, step 2
           con    cov min.nec
D + A*B   0.808  0.966     YES
-------------------------------
asf, end
```

Figure 1: Transcripts of how CNA finds **MSC** and **ASF** for outcomes C, D, and E in Table 2(c). $con_t = 0.8$ and $cov_t = 0.9$. "YES" indicates that a condition meets con_t and cov_t, respectively. The parts above the double lines report the search for **MSC**, the parts below the search for **ASF**.

data deficiencies. In configurational causal modeling, however, the need for rigorous method evaluation has not been properly recognized for too long. Only recently, a number of studies (e.g. Lucas and Szatrowski 2014; Krogslund, Choi, and Poertner 2015) criticizing QCA as an incorrect method on the basis of data simulations forcefully brought this topic onto the research agenda of configurational causal modeling. Some proponents of QCA (e.g. Olsen 2014; Ragin 2014) reacted by denying that the correctness of QCA can be tested on the basis of simulated data because the model building process of QCA allegedly requires extended case knowledge, which is unavailable in simulated data. Others (e.g. Baumgartner and Thiem 2017) concurred with the critics that QCA's correctness can be evaluated by standard benchmark tests involving simulated data but continued to produce their own simulations showing that at least the so-called *parsimonious* search strategy of QCA not only preforms faultlessly under ideal data situations but even when facing some (mild) forms of data deficiencies.

We agree with those on the latter side of this debate. Of course, case knowledge may be very helpful for CCM studies, for instance, when it comes to variable selection, calibration, model interpretation or disambiguation, but both QCA and CNA mechanically build models by means of algorithms that have no input parameters for case knowledge. Since correctness tests involving data simulations only assess a method's model building routine, they are safely applicable to both QCA and CNA. What is more, a data-driven correctness test for any method only generates conclusive results if the method cannot be immunized against failed trials. To satisfy this prerequisite, the test design must guarantee that the background assumptions of the scrutinized method are met, which, in turn, is only possible if the data-generating structure Δ is known. In (non-simulated) real-life discovery contexts, however, Δ is unknown and compliance with background assumptions can only be rendered plausible via heuristics but never ascertained beyond doubt. Consequently, when applied to real-life data, it will often be impossible to determine whether the method has actually recovered Δ; and if Δ is known but the method fails to find it, this failure cannot conclusively be ascribed to flaws in the method, as it might just as well be due to deficiencies in the data. In real-life discovery contexts, any procedure of causal inference can thus always be immunized against causal fallacies. Therefore, conclusive correctness tests must draw on simulated data.

In that light, this section reports the results of a whole battery of evaluation tests that follow the template of so-called *inverse search trials*, which reverse the order of causal discovery as it is normally conducted in scientific practice. An inverse search comprises three steps: (1) a causal structure Δ is presupposed, (2) artificial data δ is generated by letting the involved factors behave in accordance with Δ, and (3) δ is processed by a scrutinized procedure. The procedure successfully completes the inverse search iff its conclusions are

true of Δ. Inverse searches allow for rigorous correctness testing because procedures cannot be immunized against unsuccessful trials. Based on the presupposed Δ, it is possible to simulate *ideal* data—data that are free of all deficiencies of their real-life cousins. Against such an idealized background, a failure to completely recover Δ can indubitably be blamed on the procedure. In addition, it is possible to simulate non-ideal data, which allow for exploring to what degree a gradual deterioration in the data quality affects the quality of the procedure's output.

In what follows, we not only evaluate the performance of the generalized CNA algorithm introduced in the previous section, but also compare it with QCA's most reliable search strategy, viz. the parsimonious one (Baumgartner and Thiem 2017). To secure the comparability with QCA, the evaluation focuses on CNA's stages 1-3, which search for single-outcome structures (as does QCA) and constitute the method's analytical core. If CNA commits causal fallacies, these occur during the identification of atomic solutions formulas (ASF) in stages 1-3. If the sets $\mathbf{ASF}_{O_i=\omega_k}$ correctly reflect parts of the data-generating structure, the mere combination of their elements to complex solution formulas in stage 4 is straightforward and does not risk to introduce new fallacies.

For the test series, we use the R-packages cna (Ambuehl and Baumgartner 2017), which—in its newest version 2.0—implements the generalized CNA algorithm developed in this paper, and QCApro (Thiem 2016), which is the most dependable QCA software currently available. The command line interfaces of these R-packages facilitate performing inverse searches and permit the exact replication of all results—as detailed in the appended replication script. The two packages provide all functions needed for a wide array of inverse search trials. The most relevant among these functions are `randomDGS`, which randomly draws data-generating structures Δ from a factor frame $\mathbf{F}$, `allCombs`, which generates the whole space of logically possible configurations of the factors in $\mathbf{F}$, `some` and `sample`, which randomly sample a specified number of cases from a data set, `makeFuzzy`, which fuzzifies the data (e.g. to simulate background noise), `selectCases`, which selects cases that comply with Δ such that specified con_t and cov_t thresholds are satisfied, `submodels`, which generates the set of models that reflect Δ in accordance with **CC**, and `cna` and `eQMC`, which analyze the data by means of CNA and QCA, respectively.[13]

Against that background, inverse search trials as implemented in R then revolve around the following steps.

[13]For details on the parameters and arguments of these functions as well as their usage, the reader is referred to the documentations of cna and QCApro.

1. Use `randomDGS` to draw a data-generating structure Δ from a factor frame $\mathbf{F}$.
2. Use `allCombs` to generate the space α of all logically possible configurations from a factor frame $\mathbf{F}' \supseteq \mathbf{F}$.
3. If the data shall be of type fs (e.g. featuring background noise), use `makeFuzzy` to fuzzify α.
4. Use `selectCases` to select, from α, the set of cases δ complying with Δ such that Δ meets con_t and cov_t in δ.
5. If the data shall be fragmented, use `some` or `sample` to randomly sample a set of cases δ' from δ; otherwise $\delta' = \delta$.
6. If relevant factors shall be omitted from the data, eliminate columns from δ'; otherwise $\delta' = \delta$.
7. Analyze δ' by means of `cna` and `eQMC` at consistency and coverage thresholds of con_t and cov_t.
8. Check whether the outputs of `cna` and `eQMC` feature a correctness-preserving model contained in `submodels`(Δ). The trial counts as passed iff this check is positive.

Depending on the concrete data scenario to be simulated, the particularities and the arrangement of these steps must be suitably varied. More specifically, in order to simulate model *overspecification*, that is, the inclusion of factors in the simulated data that are causally irrelevant in the targeted structure Δ, $\mathbf{F}'$ must be determined to be a superset of $\mathbf{F}$ in step 2. Correspondingly, to simulate model *underspecification*, that is, the omission of factors from the data that are causally relevant in Δ, step 6 must be executed. To simulate *data fragmentation* (or limited diversity), the number of cases drawn in step 5 must be smaller than the exhaustive set of cases compatible with Δ. To simulate *inconsistencies* and *imperfect solution coverages*, con_t and cov_t must be set to values below 1 in step 4. The resulting data is of type fs if step 3 is executed, otherwise it is of type cs or mv, depending on what types of factors are chosen for Δ. Finally, to simulate data that are *ideal* by the standards of configurational causal modeling, $\mathbf{F}'$ must be identical to $\mathbf{F}$ in step 2, con_t and cov_t must be set to 1 in step 4, and steps 5 and 6 must not be executed.

We perform a total of 48 different types of tests. In each test type, we randomly draw 30 to 50 data-generating structures (depending on the calculative complexity of the analysis), on which we then perform inverse search trials using both CNA and QCA. 16 of the test types are run on cs data, 16 on fs data, and 16 on mv data. We simulate data scenarios resulting from all logically possible combinations of the following four types of data deficiencies:

overspecification (O), underspecification (U), data fragmentation (F) and imperfect solution consistencies and coverages (I). For instance, a scenario as OuFi is one *with* overspecification, *without* underspecification, *with* data fragmentation, and *without* imperfect (i.e. with prefect) consistencies and coverages, OUFI, in contrast, features all four types of data deficiencies, while oufi is free of all deficiencies and, hence, results in ideal configurational data.

In order to keep the whole test series easily replicable, the complexity of the randomly drawn data-generating structures is kept comparably simple: they feature between three and four exogenous factors and one outcome each. To simulate overspecification, one irrelevant factor is added to the data; and underspecification is simulated by removing one relevant factor from the data. Moreover, in scenarios with data fragmentation, half of the cases that are compatible with the data-generating structure are removed in case of cs or fs data, while in case of mv data we remove 80% of the compatible cases; that is, we simulate diversity indices of 0.5 and 0.2, respectively. Finally, in scenarios with imperfect solution consistencies and coverages, the targeted data-generating structures are set to only reach consistencies and coverages of 0.8 in the simulated data.

Table 3 lists the correctness ratios obtained in each test type, that is, the ratios of the number of trials complying with configurational correctness (**CC**) to the total number trials in each test type. For instance, "1" means that every trial satisfied **CC** or "0.77" that 77% of the trials did. A number of aspects of the results in Table 3 deserve separate emphasis. First, CNA significantly outperforms QCA in regard to correctness in a number of data scenarios and performs equally well in all other scenarios. Second, ideal configurational data of all three data types are faultlessly analyzed by both methods.[14] Even in scenarios with certain data deficiencies, both methods reach perfect correctness ratios. In cs and fs data, this holds in particular when overspecification and data fragmentation are not complemented by either underspecification or imperfect consistencies and coverages. In mv data, even combinations of over- and underspecification do not diminish correctness ratios. Third, as is to be expected, neither method performs without error in the increasingly deficient data scenarios. No method can faultlessly analyze deficient data that do not faithfully reflect data-generating structures. But while QCA's correctness ratios plummet in certain cases, in particular, when over- and underspecification are combined with imperfect consistencies and coverages, CNA maintains reasonable correctness ratios even in those cases.

[14] In regard to the evaluation of QCA, this finding confirms recent results of Baumgartner and Thiem (2017) and contrasts with claims made by Lucas and Szatrowski (2014).

test type	#	cs data		#	fs data		#	mv data	
		CNA	QCA		CNA	QCA		CNA	QCA
oufi	1	1	1	17	1	1	33	1	1
Oufi	2	1	1	18	1	1	34	1	1
oUfi	3	1	0.86	19	0.9	0.8	35	1	1
ouFi	4	1	1	20	1	1	36	1	1
oufI	5	0.94	0.66	21	0.98	0.8	37	1	0.22
OUfi	6	0.77	0.6	22	0.9	0.83	38	1	1
OuFi	7	1	1	23	1	1	39	1	1
oUFi	8	0.97	0.97	24	0.93	0.9	40	0.83	0.57
ouFI	9	0.9	0.84	25	0.84	0.7	41	0.86	0.8
OufI	10	0.94	0.16	26	0.8	0.34	42	1	0.12
oUfI	11	1	0.7	27	0.87	0.7	43	1	0.97
OUFi	12	0.8	0.43	28	0.9	0.87	44	0.83	0.1
OUfI	13	0.83	0.1	29	0.9	0.4	45	0.93	0.27
OuFI	14	0.8	0.58	30	0.66	0.36	46	0.88	0.86
oUFI	15	0.9	0.53	31	0.8	0.77	47	0.73	0.17
OUFI	16	0.7	0	32	0.8	0.77	48	0.8	0.03

Table 3: Correctness ratios for each test type. The tests are numbered (#) in correspondence with the replication script.

A proper interpretation of this last finding requires some differentiation. In particular, it must not be misinterpreted to mean that CNA could always squeeze more reliable information out of deficient data than QCA. The CNA output is only more reliably informative in scenarios with combinations of overspecification and imperfect consistencies and coverages but without underdetermination, that is, in scenarios of types OufI and OuFI. In those scenarios, CNA's bottom-up approach allows the method to reliably abstain from an erroneous causal interpretation of irrelevant factors, where QCA tends to fail at eliminating the latter because it builds models from the top down. By contrast, in data scenarios featuring over- and underspecification in combination with further deficiencies, CNA's edge over QCA is not so much due to its capacity to distill more adequate information out of deficient data but rather to the fact that CNA is more cautious than QCA in drawing causal inferences. CNA imposes both consistency and coverage thresholds on the causal models, to the effect that, if these thresholds cannot be met because of data deficiencies, CNA abstains from issuing any causal models and, thus, does not commit causal fallacies. By contrast, as we have

seen in section 2.3, QCA uses neither consistency nor coverage thresholds as binding model building criteria; instead, it merely imposes consistency cutoffs on sufficient conditions in the course of truth table generation. Consequently, QCA continues to draw causal inferences where CNA abstains from doing so. In cases of severely deficient data this leads to QCA committing causal fallacies where CNA issues no models.

In general, a method of causal inference must not only aim to distill as much information out of the processed data as possible, but it must also have built-in controls that prevent causal inferences when the data quality is insufficient. Hence, a correctness ratio as 0.9 in, say, test #27 does not imply that the CNA output comprises a (non-empty) correct model in 9 out of 10 trials, rather it implies that CNA does not commit a causal fallacy in 9 out of 10 trials—which, in case of #27, is mostly realized by CNA issuing no model at all and, thereby, indicating to the analyst that there is something wrong with the data. QCA also has some such in-built controls; for instance, towards the lower end of the fs test series, QCA frequently aborts the analysis with error messages, which leads to QCA's correctness ratio increasing despite decreasing data quality. But, in its current state, QCA does not have enough of these controls. There are too many scenarios of data deficiencies that QCA fails to recognize as such. Most of all, the test series conducted here indicates that QCA would urgently need a filter that imposes both consistency and coverage thresholds on the issued causal models.

To round off this evaluation, we not only culled correctness ratios from the 48 test types, but also completeness ratios. The completeness ratio in a test type corresponds to the ratio of the number of trials in which a method *completely* uncovers the data-generating structure to the total number of trials. Completeness ratios are reported in Table 4. As is to be expected, CNA and QCA can only systematically uncover all properties of data-generating structures when the data quality is very high. Moreover, it is clear that in cases of underspecification neither method has a chance of ever finding the complete data-generating structure. Apart from these confirmations of theoretical expectations, Table 4 shows that the completeness ratios of the two methods are very close together across the whole test series, except for the tests #10, #21, #26, #37, and #42 where CNA has a significant edge over QCA. That is, CNA's superior correctness ratios are not offset by lower completeness ratios; rather, with regard to completeness, CNA likewise outperforms QCA in a number of data scenarios and performs comparably in all other scenarios.

Let us end this discussion with two qualifications. First, neither correctness nor completeness ratios reflect ambiguities ratios, that is, the number of equally viable models returned in each trial. As indicated in section 2.4, configurational data often underdetermine

test type	#	cs data		#	fs data		#	mv data	
		CNA	QCA		CNA	QCA		CNA	QCA
oufi	1	1	1	17	1	0.94	33	1	1
Oufi	2	1	1	18	1	0.9	34	1	1
oUfi	3	0	0	19	0	0	35	0	0
ouFi	4	0.06	0.06	20	0.08	0.12	36	0.16	0.16
oufI	5	0.46	0.42	21	0.7	0.16	37	0.84	0.16
OUfi	6	0	0	22	0	0	38	0	0
OuFi	7	0.34	0.34	23	0.5	0.48	39	0.48	0.48
oUFi	8	0	0	24	0	0	40	0	0
ouFI	9	0.02	0.02	25	0.14	0.04	41	0.1	0.1
OufI	10	0.68	0.12	26	0.66	0	42	0.86	0.12
oUfI	11	0	0	27	0	0	43	0	0
OUFi	12	0	0	28	0	0	44	0	0
OUfI	13	0	0	29	0	0	45	0	0
OuFI	14	0.14	0.18	30	0.16	0.02	46	0.08	0.08
oUFI	15	0	0	31	0	0	47	0	0
OUFI	16	0	0	32	0	0	48	0	0

Table 4: Completeness ratios for each test type. The tests are numbered (#) in correspondence with the replication script.

their own causal modeling. In consequence, both CNA and QCA frequently issue considerably more than one causal model. Hence, even in a trial in which both CNA and QCA correctly and completely identify the data-generating structure, the user might be presented with numerous models, only one of which is the targeted structure. None of the results presented here have any implications for CNA's and QCA's ambiguities ratios (which must be scrutinized in a future test series). Second, the forms of data deficiencies analyzed here do not exhaust the space of possible deficiencies. For instance, all the data simulated here feature evenly distributed case frequencies, that is, different configurations are represented by equally many cases. Of course, that is often not the case in real-life data. It is thus an open question how CNA and QCA fare and compare under biased case frequencies. Also, the test series conducted in this paper sets consistency and coverage thresholds and diversity indices to constant values without exploring how the two methods perform under variations of these values.

5 Conclusion

This paper has generalized Coincidence Analysis (CNA), a configurational comparative method of causal data analysis, for multi-value variables and variables with continuous values from the unit interval that are interpreted as membership scores in fuzzy-sets. Moreover, it has shown in an extended series of benchmark tests that CNA performs both correctly and completely in ideal data scenarios and maintains reliable correctness ratios across a wide range of data deficiencies.

CNA differs from QCA, the currently dominant CCM, in numerous respects. First, CNA not only uncovers single-outcome structures but also structures with multiple outcomes. It is the only CCM custom-built to uncover the Boolean complexity dimension of sequentiality. Second, CNA builds causal models from the bottom up rather than from the top down. Thereby, it renders redundancy elimination (or minimization) itself redundant—which constitutes the algorithmic core of QCA. This reversal of the basic model building approach, on the one hand, allows CNA to abstain from erroneously causally interpreting irrelevant factors in cases of model overspecification and, on the other, permits CNA to directly apply one and the same algorithmic protocol to all data types, without a detour via truth tables. Third, CNA imposes authoritative consistency and coverage cutoffs on causal models (and all their elements), whereas QCA only uses a consistency threshold in truth table generation. In consequence, CNA is much more risk-averse than QCA when it comes to drawing causal inferences, which, in turn, yields that CNA maintains reasonably high correctness ratios even in scenarios featuring severe data deficiencies that cause QCA's ratios to plummet. At the same time, we have seen that this inferential caution does not entail that CNA would fail to completely uncover data-generating structures where QCA succeeds in doing so.

Overall, the generalized version of CNA not only reliably uncovers all Boolean dimensions of causal structures from crisp-set, multi-value, and fuzzy-set data, but also has effective inbuilt controls that abandon an analysis that is too risky due to data deficiencies. In that light, CNA constitutes a powerful methodological alternative for researchers interested in the Boolean dimensions of causality.

References

Ambuehl, M. and M. Baumgartner (2017). *cna: Causal modeling with Coincidence Analysis*. URL: https://cran.r-project.org/package=cna.

Baumgartner, M. (2009a). Inferring causal complexity. *Sociological Methods & Research 38*, 71–101.

Baumgartner, M. (2009b). Uncovering deterministic causal structures: A Boolean approach. *Synthese 170*, 71–96.

Baumgartner, M. (2014). Parsimony and causality. *Quality & Quantity*, 1–18.

Baumgartner, M. and A. Thiem (2015). Model ambiguities in configurational comparative research. *Sociological Methods & Research*. doi: 10.1177/0049124115610351.

Baumgartner, M. and A. Thiem (2017). Often trusted but never (properly) tested: Evaluating Qualitative Comparative Analysis. *Sociological Methods & Research*. doi: 10.1177/0049124117701487.

Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

Cronqvist, L. (2017). *Tosmana: Tool for Small-N Analysis [Computer Programme], Version 1.53*. Trier: University of Trier. url: http://www.tosmana.net.

Cronqvist, L. and D. Berg-Schlosser (2009). Multi-value QCA (mvQCA). In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques*, pp. 69–86. London: Sage Publications.

Downing, B. M. (1992). *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. N.J.: Princeton University Press.

Duşa, A. (2007). User manual for the QCA(GUI) package in R. *Journal of Business Research 60*(5), 576–586.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Goertz, G. (2006). *Social Science Concepts: A User's Guide*. Princeton: Princeton University Press.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

Hájek, P. (1998). *Metamatematics of Fuzzy Logic*. Dordrecht: Kluwer.

Hume, D. (1999 (1748)). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.

Krogslund, C., D. D. Choi, and M. Poertner (2015). Fuzzy sets on shaky ground: Parameter sensitivity and confirmation bias in fsQCA. *Political Analysis 23*(1), 21–41.

Lucas, S. R. and A. Szatrowski (2014). Qualitative Comparative Analysis in critical perspective. *Sociological Methodology 44*(1), 1–79.

Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.

Mill, J. S. (1843). *A System of Logic*. London: John W. Parker.

Olsen, W. (2014). Comment: The usefulness of QCA under realist assumptions. *Sociological Methodology 44*(1), 101–107.

Ragin, C. C. (1987). *The Comparative Method*. Berkeley: University of California 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. (2009). Qualitative Comparative Analysis using fuzzy sets (fsQCA). In B. Rihoux and C. C. Ragin (Eds.), *Configurational Comparative Methods*, pp. 87–121. Los Angeles: Sage.

Ragin, C. C. (2014). Comment: Lucas and Szatrowski in critical perspective. *Sociological Methodology 44*(1), 80–94.

Ragin, C. C. and S. Davey (2016). *fs/QCA: Fuzzy-set/Qualitative Comparative Analysis, version 3.0 [computer program]*. Irvine: University of California.

Schneider, C. Q. and C. Wagemann (2010). Standards of good practice in Qualitative Comparative Analysis (QCA) and fuzzy-sets. *Comparative Sociology 9*, 397–418.

Simon, H. A. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association 49*(267), 467–479.

Spirtes, P., C. N. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2nd ed.). Cambridge, MA: MIT Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North Holland.

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

Thiem, A. (2016). *QCApro: Professional Functionality for Performing and Evaluating Qualitative Comparative Analysis, R Package Version 1.1-0*. URL: https://cran.r-project.org/package=QCApro.

Thiem, A., M. Baumgartner, and D. Bol (2016). Still lost in translation! A correction of three misunderstandings between configurational comparativists and regressional analysts. *Comparative Political Studies 49*, 742–774.

Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. New York: Oxford University Press.