Necessary Condition Analysis (NCA):
Logic and methodology of "necessary but not sufficient" causality

Jan Dul

Rotterdam School of Management, Erasmus University Rotterdam, the Netherlands.

Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands. E-mail: jdul@rsm.nl. Tel: +31-10-408-1719

Abstract

Theoretical "necessary but not sufficient" statements are common in the organizational sciences. Traditional data analyses approaches (e.g., correlation or multiple regression) are not appropriate for testing or inducing such statements. This paper proposes Necessary Condition Analysis (NCA) as a general and straightforward methodology for identifying necessary conditions in datasets. The paper presents the logic and methodology of necessary but not sufficient contributions of organizational determinants (e.g., events, characteristics, resources, efforts) to a desired outcome (e.g., good performance). A necessary determinant must be present for achieving an outcome, but its presence is not sufficient to obtain that outcome. Without the necessary condition, there is guaranteed failure, which cannot be compensated by other determinants of the outcome. This logic and its related methodology are fundamentally different from the traditional sufficiency-based logic and methodology. Practical recommendations and free software are offered to support researchers to apply NCA.

Keywords: necessity, sufficiency, multi-causality, data analysis, software

According to David Hume's philosophy of causation: "...*we may define a cause to be an object, followed by another, and where all the objects, similar to the first, are followed by objects, similar to the second. Or in other words, where, if the first had not been, the second never had existed.*" (Hume, 1777). This notion of causality can be interpreted as having two distinct logical parts (Abbott, 1974; Goertz, 2003; Lewis, 1973; Pizzi, 1993). The first quoted sentence reflects the common understanding of causality, namely that X is sufficient for Y: "if X, then Y". Most people tend to understand "X causes Y" in terms of sufficiency (Mandel & Lehman, 1998; Wasserman, Dorner, & Kao, 1990). This may be the reason that hypotheses in organizational sciences frequently imply sufficiency. The second quoted sentence reflects that X is necessary for Y: "no Y without X". Such hypotheses are rarely formulated and tested in organizational sciences (for an exception see McLachlin, 1997 as discussed in Dul, Hak, Goertz, & Voss, 2010).

Although scholars often confuse necessity and sufficiency (Chung, 1969; Goertz & Starr, 2003), the two are totally different[1]. A necessary cause allows an outcome to exist; without the necessary cause, the outcome will not exist. A sufficient cause ensures that the outcome exists; it produces the outcome. A student who wants to be admitted to a USA Graduate School (the outcome) needs to have a high score on the Graduate Record Examinations (GRE) test: an adequate GRE score is necessary for the outcome. Necessary causes are not automatically sufficient. An adequate GRE score is not sufficient for admission because also other admission requirements play a role (e.g., the student's motivation letter, a good TOEFL score, reputation of the student's bachelor program, recommendation letter). However, if the student's GRE score is too low, there is guaranteed failure, independently of the student's performance on the other requirements. Therefore, a necessary cause is a constraint, a barrier, an obstacle, a bottleneck, that must be managed to allow a desired outcome to exist. Each single necessary cause must be in place, as there is no additive causality that can compensate for the absence of the necessary cause. Prevention of guaranteed failure, and increased probability of success are core constituents of the "necessary but not sufficient" logic of "X causes Y"[2].

This paper presents a new family of analytical approach that may be used to test or induce hypotheses examining the necessary, but not sufficient, contributions of various organizational determinants (e.g., events, characteristics, resources, efforts) to various outcomes (e.g., individual job attitudes, firm performance). To be clear, this paper is about a new data analytic tool, denoted Necessary Condition Analysis or NCA, which may be used alongside traditional data analytic tools (e.g., multiple regression, structural equation model).

Both traditional data analysis and NCA may be used to draw causal inferences. However, the validity of any causal inference is predicated on the adequacy of theory, the quality of measurement, and the particular research design used by researchers (cf. James, Mulaik, & Brett, 1982; Mathieu & Taylor, 2006; Shadish, Cook, & Campbell, 2002; Stone-Romero, 2002). Consequently, NCA (just as traditional analytical tools) is not immune to problematic theory (e.g., omitted variables problem), measurement (e.g., use of unreliable measures), or design (e.g., use of a weak research design). With that important caveat in place, we may proceed with a comparison of traditional approaches with NCA. Traditional paradigms of multi-causality of outcomes presume that each cause is sufficient to increase the outcome, but none is necessary. In such paradigms, causality is additive. Determinants add up to cause the outcome. This can be expressed as the additive model $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots$, which is common in multiple linear regression. If one determinant becomes zero, the outcome will be reduced, and this effect can be compensated by increasing the values of other determinants. In *additive* Graduate school admission logic, this would mean that a GRE score below the required level can be compensated for by having a higher TOEFL score. In the "necessary but not sufficient" paradigm, absence of the necessary determinant immediately results in outcome failure, independently of the value of the other determinants. In *necessary condition* Graduate school admission logic, this means that if the GRE score is below the required level, it makes no sense to work hard on improving the TOEFL score. Necessary causality may be better expressed as a multiplicative phenomenon: $Y = X_1*X_2*X_3\ldots$ (Goertz, 2003)[3]. The dramatic sudden effect of zero values for necessary conditions fits many everyday experiences. A car stops moving if the fuel tank is empty; financial markets collapse if trust is gone. These bottlenecks must be taken away first, before the desired outcome can be achieved. In these examples, fuel and trust are necessary but not sufficient for the outcome.

In the social sciences in general, and in the organizational sciences in particular, theoretical statements in terms of "necessary but not sufficient" exist widely. This may be because necessary determinants have great practical relevance and impact. Absence of the necessary determinants prevents the organization from better performance. Table 1 presents results of a search for "necessary but not sufficient" and "necessary but insufficient" statements in a leading organizational science journal (Academy of Management Journal). In research papers such statements are commonly made when developing hypotheses (e.g., in the introduction or a theory development section) or when evaluating theoretical or practical consequences of the empirical test results (e.g., in the discussion section). The table shows that researchers use traditional additive causality models and related data analysis approaches

(e.g., correlation, regression) for empirically testing or inducing necessary but not sufficient statements. I would like to add one important caveat with respect to the papers included in Table 1. The authors of these papers are pursuing relevant research questions; unfortunately, the existing library of data analytic tools was not specifically developed to address these types of questions. Consequently, the authors were restricted to using traditional analytical tools. Thus, my criticism is not directed at the authors or how they approached their data analysis. My criticism is related to the lack of tools that were available to these authors. One of the goals of the current paper is to introduce a new set of analytical tools that may be of value to researchers pursuing questions such as those summarized in Table 1.

With that caveat in place, many of the papers presented in Table 1 (and identified during my review but not included in Table 1) tend to adhere to the following methodological sequence:

1. A theoretical necessary but not sufficient statement is introduced in the Introduction, Theory, or Hypotheses section.

2. This statement is reformulated as a traditional hypothesis with an unspecified "positive association" (or similar) between determinant and outcome (i.e., without specifying that the determinant is considered as necessary but not sufficient).

3. The obtained observational data are analyzed using traditional variants of the general linear model (e.g., correlation or regression).

4. If the hypothesis is supported (e.g., a positive association found between determinant and effect), the results are interpreted as a support of the necessary but not sufficient statement.

Table 1

Examples of "Necessary But Not Sufficient" Statements in Papers Published in the Academy of Management Journal (emphases not in original;

Bold*: the assumed causal relationship;* Underlined*: the determinant X; Italics: the outcome Y).*

Source	Research Design and Sample	Theoretical "necessary but not sufficient statement"	Hypothesis	Data Analysis	Interpretation
Autio, Dahlander, & Frederiksen (2013)	Cross-sectional observational study of 275 members of an online user community	"The recognition of third-person entrepreneurial opportunities—that is, the conjecture that an opportunity exists for the willing and able to sell products and services at a price greater than the cost of their production—is a **necessary but not sufficient** prerequisite for *entrepreneurial action.*"	"We expect **a correlation** between opportunity evaluation and *entrepreneurial action.*"	Correlation, Visual inspection of 2x2 matrix	"*Entrepreneurial action* **was more likely** among individuals who engaged in opportunity evaluation". "We did not treat opportunity evaluation as a **necessary condition** for *entrepreneurial action.*"
Diestre & Rajagopalan (2011)	Cross-sectional observational study of 1171 firms in manufacturing industries	"A greater exposure to the chemical compounds that a firm already uses will **allow it** to reach higher levels on the learning curves associated with those materials, and this knowledge should improve its *competitive positions in both the new industry and its home industry.* In other words, firms will **be more likely** to *enter chemically related industries* as a way to exploit current resources and capabilities and accumulate additional expertise in using similar chemicals."	"The level of chemical relatedness between a firm's home industry is **positively associated** with the likelihood that a firm will *diversify into that target industry.*"	Multiple regression (Logit)	"Our study shows that the opportunity to exploit sources of relatedness between different industries is a **necessary but not sufficient condition** for a firm *to diversify.*"
Kirca, et al. (2011)	Meta-analysis of 120 samples (of mainly cross-sectional observational studies)	"The existence of these assets constitutes a **necessary but not sufficient condition** for their *cross-border transfer within an MNE.*"	"the level of firm-specific assets *is* **positively related** to *firm multinationality.*"	Correlation, Multiple regression	"R&D intensity and advertising intensity have **positive effects** on *firm multinationality.*"
Rubin, Munz, & Bommer (2005	Cross-sectional observational study of 145 leaders and their subordinates in one company	"Since leader-follower relationships are **critical** to successful leadership, and the ability to recognize emotion **is important** for building strong leader-follower relationships, it stands to reason that *transformational leaders* who build strong relationships do so in part through understanding followers' emotions"	"leader emotion recognition ability *is* **positively associated** with leader *transformational leadership behavior.*"	Multiple regression	"[o]ur results suggest that emotion recognition may be a **necessary but insufficient** ability involved in the performance of *transformational leadership behavior.*"
Martin & Eisenhardt (2010)	Multiple case study with 12 cases		"how social relationships amongst Business Units	Qualitative analysis;	"Social relationships are **necessary but not sufficient** for

			(and their managers) within firms **influence** *cross-BU collaborations."* "Rewiring – a business-unit-centric process **is more likely to yield** *a high- performing collaboration* than is a corporate-centric process"	visual inspection of data matrices	promoting *high-performing cross-BU collaboration."*
(collaborations) from 6 firms (theory building)					
Hausknecht, Hiller, & Vance (2008)	5 waves of data collection in 115 work units of one organization	"When the level of analysis is the work unit, <u>organizational commitment</u> may be a **necessary but insufficient condition** for low *absenteeism."*	"when <u>organizational commitment</u> is low, **higher levels of** *absenteeism* **may occur**."	Multiple regression	"[…] findings […] show that high <u>organizational commitment</u> is **necessary** to avoid high levels of *absenteeism."*

Consider a few illustrative examples from Table 1. Kirca et al. (2011) studied firm specific assets as a cause of multinationality of international firms (which is a driver of performance of such firms) and stated in their Theory and Hypotheses section that R&D intensity and advertising intensity are firm specific assets that "constitute a *necessary but not sufficient* condition for their cross border transfer within an MNE" (p. 50, italics added). Subsequently, they re-formulated this relationship as "the level of firm-specific assets is *positively related* to firm multinationality" (p. 50, italics added), and tested and confirmed this hypothesis using traditional correlation and regression analyses. In another example, Rubin, Munz, and Bommer (2005) studied the effect on transformational leadership behavior of a leader's ability to recognize a follower's emotion, and formulated the hypothesis that "leader emotion recognition ability is *positively associated* with leader transformational leadership behavior" (p.848, italics added). Using regression analysis, they found support for this hypothesis and concluded in the Discussion section that "[o]ur results suggest that emotion recognition may be a *necessary but insufficient* ability involved in the performance of transformational leadership behavior". (p.854, italics added). In yet another example, Hausknecht, Hiller, and Vance (2008) studied effects of organizational commitment on work-unit absenteeism and claimed in their Theory and Hypotheses section that "organizational commitment may be a *necessary but not sufficient* condition for low absenteeism". But then they re-formulated the statement in sufficiency terms as "when organizational commitment is low, higher levels of absenteeism *may occur.*" (p. 1226, italics added), tested it by using traditional regression analysis, and claimed in the Discussion section that "[…] findings […] show that high organizational commitment *is necessary* to avoid high levels of absenteeism" (p.1239, italics added).

In each of the above examples, the authors were restricted in how they could go about testing or inducing a hypothesis framed in terms of "necessary but not sufficient" relationships. The above examples are simply illustrative, and are not unique to this set of authors or this literature. Instead, they demonstrate how (across a wide range of management topics) researchers seek to test or induce necessary but not sufficient hypotheses, but lack the proper analytical tools for doing so. Researchers are often unaware of the importance, the special characteristics, and the required methodology of the necessary but not sufficient approach. In this paper, such an alternative data analysis approach is presented. Instead of drawing trend lines "through the middle of the data" in scatterplots (such as regression lines), it searches for empty areas in scatterplots and draws "ceiling lines" that separate empty and

full data areas (for a detailed discussion on ceiling lines see subsection Ceiling Techniques). I will distinguish between three types of necessary conditions: dichotomous, discrete, and continuous, and will use three illustrative examples to explain their similarities and differences: I will extend on the above *dichotomous* example of necessary GRE scores for admission (Vaisey, 2009), present a *discrete* example from Table 1 about necessary representation in a collaboration team for successful collaboration (Martin and Eisenhardt, 2010), and discuss a *continuous* example about necessary personality traits for sales performance (Hogan & Hogan, 2007). I integrate earlier work and ideas on analyzing dichotomous necessary conditions (e.g., Braumoeller & Goertz, 2000; Dul et al., 2010), discrete necessary conditions (e.g., Dul et al., 2010), and continuous necessary conditions (e.g., Goertz, 2003; Goertz, Hak, & Dul, 2013) and combine these with new ideas e.g., "multivariate necessary condition", "bottleneck technique", "necessity inefficiency", "effect size" of a necessary conditions, and techniques for drawing ceiling lines. I also present practical recommendations and software to guide researchers to perform an NCA analysis.

I consider NCA as a complement, not a replacement of traditional approaches to analyze causal relations. NCA may provide new insights that are normally not discovered with traditional approaches: Multiple regression may spot determinants that (on average) contribute to the outcome (i.e., determinants with large regression coefficients), whereas NCA may spot necessary (critical) determinants that prevent an outcome to occur ("constraints" or "bottlenecks"). When bottlenecks are present, performance will not improve by increasing the values of other determinants (e.g., determinants with large regression coefficients), unless the bottleneck is taken away first (e.g., by increasing the value of the critical determinant). Only in those situations when none of the determinants are critical for reaching a desired outcome can traditional approaches inform us how to increase performance. Therefore, NCA may precede traditional approaches. This necessary but not sufficient approach is needed because critical determinants may or may not be part of lists of determinants identified with traditional approaches. Furthermore, due to the fundamental difference between the two approaches, determinants that show small or zero effects in traditional approaches may be identified as critical in the necessary but not sufficient approach (e.g., see the example presented in Figure 4), and determinants that show large effects in traditional approaches may not be critical according to the necessary but not sufficient approach. Indeed, both approaches are essential for a proper understanding of organizational phenomena. The NCA approach also adds to insights obtained from Qualitative Comparative Analysis (QCA) (Ragin, 2000, 2008), which is another approach that understands causality in terms of necessity and sufficiency, rather

than correlation or regression. NCA focuses on (levels of) single determinants (and their combinations) that are necessary but not automatically sufficient, whereas QCA focuses on combinations of determinants (configurations) that are sufficient but not automatically necessary. Although QCA centers on sufficient configurations it can also analyze single necessary conditions, but it does so in a limited way (for a further discussion on the difference between NCA and QCA see subsection Comparison NCA with QCA).

In what follows, I first discuss the logic of necessary conditions in more detail. Next, I present the NCA methodology by discussing ceiling techniques, effect size of the necessary condition, "necessity inefficiency", the comparison of NCA with QCA, and limitations of NCA. Finally, I offer practical recommendations and free software for applying NCA.

The Logic of Necessary Conditions

In the above examples, both the condition (GRE-score, fuel, trust) and the outcome (admission, moving car, solid financial system) can have only two values: absent or present. This is the dichotomous logic of necessary conditions, which is part of classic binary (or in mathematics, Boolean) logic. However, many real life situations are not inherently dichotomous (although they can be dichotomized). In the discrete situation, the necessary condition and the outcome can have more than two values (e.g., low, medium, high), and in the continuous situation, any value is possible. In this section, I discuss the logic of necessary conditions by moving from the dichotomous, via the discrete, towards the continuous necessary condition.

Dichotomous Necessary Condition

Figure 1 (top) shows two ways to graphically represent the dichotomous necessary condition. On the left, possible observations for a necessary condition are plotted in the center of cells of a contingency matrix (I will use the convention that the X-axis is "horizontal" and the Y-axis is "vertical" and that values increase "upwards" and "to the right"[4]). The contingency matrix is a common way to present dichotomous necessary conditions (e.g., Braumoeller & Goertz, 2000; Dul et al., 2010). On the right, the possible observations are plotted on a grid of a Cartesian coordinate system. This representation is used here to facilitate the generalization of necessary condition characteristics from dichotomous logic towards discrete and continuous logic (see later). The dashed lines are the ceiling lines that separate the area with observations from the area without observations. The figure (top) shows that outcome $Y=1$ is only possible if $X=1$; however if $X=1$ two outcomes are possible, $Y=0$ and $Y=1$. Therefore $X=1$ is
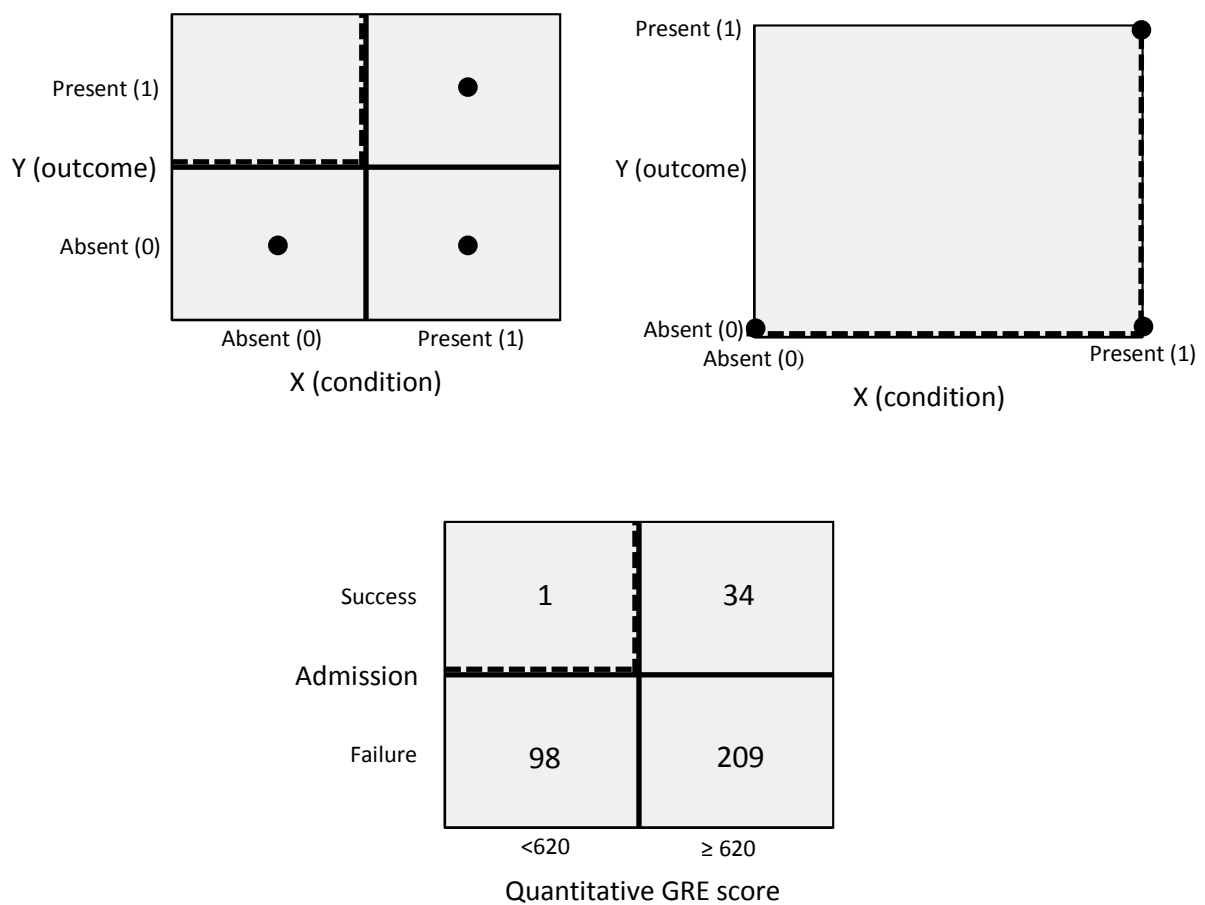
Figure 1. The dichotomous necessary condition with possible observations shown in the center of the cells of a contingency matrix (top left) and in the grid of a Cartesian coordinate system (top right). The dashed line is the "ceiling line" separating the area with and the area (virtually) without observations. At the bottom is an example of a contingency table where X is low or high Quantitative GRE score, and Y is Admission failure or success. In the center of the cells is the number of students applying for admission to the Berkeley Sociology Graduate Program in 2009 (Adapted from: Goertz & Mahoney, 2012; Vaisey, 2009).

necessary but not sufficient for Y=1. However, if X=0, then Y=0, the absence of X is sufficient for the absence of Y. The *necessity of the presence* of X for the presence of Y, and the *sufficiency of the absence* of X for the absence of Y are logically equivalent. Hume's (1777) statement mentioned in the introduction section of this paper, i.e., "*if the first had not been, the second never had existed.*" is the "sufficiency of absence" version of the necessary condition formulation.

The "sufficiency of absence" formulation of the necessary condition allows the value of X=0 to be considered as a constraint or bottleneck for achieving the high value of the outcome (Y=1). The empty cell for X=0, Y=1 indicates that a high value of Y=1 cannot be achieved for the low value of X=0. If the high Y-value is to be achieved, the bottleneck must be removed by realizing a high value of X. On the other hand, the "necessity of presence" formulation of the necessary condition refers to the general organizational challenge of putting in place the right level of scarce resources or (managerial) effort to be able to reach a desired level of organizational performance (the outcome). Then X can be considered as an organizational ingredient (scarce event/characteristic/resource/effort) that is needed, and therefore allows the desired organizational outcome Y (good performance) to occur.

The two versions of the same logic imply two possible ways of testing or inducing necessary conditions with datasets of observations (Dul et al., 2010). In the first way, different to traditional correlation or regression analyses, only successful cases are purposively sampled (for an everyday example see Appendix 1). Conversely, a correlation or regression analysis requires variance in the dependent variable, hence, then both successful and unsuccessful cases must be randomly sampled and the analysis shows the extent to which determinants on average contribute to success. In NCA, the presence or absence of the condition is determined in cases with the outcome present (successful cases: Y=1). The presence of the condition in successful cases is an indication of necessity of X=1 for Y=1[5]. If the condition is not present in the successful cases, then X is not necessary for Y. This approach was used by the current editor of Organizational Research Methods (ORM) before he took office as editor in July 2013. He performed a Google Scholar search to identify highly cited ORM papers and their common characteristics and found, for example, that such papers give practical recommendations to the organizational researcher (LeBreton, personal communication June 26, 2014). Additionally, in his first editorial, LeBreton (2014, p. 114) states that "over the past 10 years, I have spent considerable time pondering what makes an impactful methods article. I have concluded that such articles share the following characteristics…" The editor (implicitly) performed a necessary condition analysis: selecting cases purposively on the basis of the presence of the dependent variable (successful cases, i.e., highly cited papers), and identifying common determinants that were present in these cases. This resulted in necessary conditions for a highly cited paper in ORM, and several of these are listed in the editorial to stimulate authors, reviewers, and editors to produce impact papers in ORM, by ensuring that the necessary conditions are in place (although this does not guarantee

impact success). Necessary condition analysis is an effective and efficient way to find relevant variables that must be managed for success.

In the second approach to test necessary conditions, cases are selected in which the condition is absent (X=0). The absence of the outcome in cases without the condition is an indication of necessity of X for Y. If success is present in cases without the condition, X is not necessary for Y. The editor could have searched for papers without practical recommendations to organizational researchers and –if the necessary condition holds– would have found that these papers are not highly-cited. Normally, this second approach to test necessary conditions is much less efficient than the first one.

Two other approaches for selecting cases will not work. It is a common misunderstanding that necessary conditions can be detected by selecting and evaluating failure cases. There is no point in evaluating the characteristics of papers that are not highly cited. Figure 1 (top) shows that falsification of a necessary condition is not possible with failure cases (Y=0) because the condition can be both absent (X=0) and present (X=1). In a similar way, an analysis of cases where the condition is present is not insightful either. A necessary condition is characterized by having no observations in the upper left hand corner. Therefore, the analysis of necessary but not sufficient conditions is about "the search for empty space" in the upper left hand corner. This applies not only to the dichotomous case, but also to the discrete and continuous cases[6].

**Illustrative Example of the Dichotomous Necessary Condition**

Figure 1 (bottom) shows an example of Quantitative GRE scores of 342 students applying for admission to the Berkeley Sociology Graduate Program in 2009 (Goertz & Mahoney, 2012; Vaisey, 2009). A traditional data analysis to test the association between GRE score and admission would show a high correlation between the two variables (tetrachoric correlation coefficient 0.6). The data justify the conclusion that success is more likely when the GRE score is at least 620 than when it is below 620. However the necessary but not sufficient interpretation of the data is that a score of at least 620 is a virtual[7] (only one exception) necessary condition for admission. The exception is a student that was admitted based on a faculty member's explicit testimony to the student's quantitative abilities, which was regarded as superior information to the quantitative GRE score (Vaisey, personal communication July 2, 2014). The upper left hand corner of Figure 1 is almost empty. Virtually all students (99%) who scored below 620 failed to be admitted. But scoring at least 620 did not guarantee success: only 14% of these students succeeded in being admitted.

Having a score below 620 is a bottleneck for being admitted. There may be other bottlenecks as well (e.g., a TOEFL score below 100). Advising students based on a traditional analysis would be to "score high on GRE and TOEFL to increase your chances for success", whereas advising them based on a necessary but not sufficient analysis would add "and make sure that you reach *all* minimum requirements to avoid guaranteed failure, because you cannot compensate an inadequate score for one requirement with a high score for another one".

Discrete Necessary Condition

In the organizational sciences, variables are often scored by a limited number of discrete values, which usually have a meaningful order (ordinal scale), e.g., a value can be higher/more/better than another value. Additionally, the differences between ordinal values can be meaningful, e.g., the difference between high and medium may be the same as the difference between medium and low (interval scale). Many data in organizational sciences are obtained from measurements with respondents or informants using standardized questionnaires with Likert-type scales. In such an approach, a variable is usually represented by a group of items with several response categories. If interval properties are assumed, item responses can be combined (e.g., by adding or averaging) to obtain a score for the determinant of interest. If items are scored on a discrete scale, the resulting score will be discrete as well. For example, if a variable is constructed from five items, each having seven response categories from 1 to 7, the number of discrete values of the variable is 31[8].

Figure 2 shows a trichotomous example of possible observations for a discrete necessary condition, where both X and Y have three possible values: Low (0), Medium (1), and High (2). The empty area in the upper left hand corner indicates the presence of a necessary condition. In this paper, it is assumed that the ceiling line in a contingency table that separates the "empty" cells without observations from the "full" cells with observations, as well as the ceiling line in a Cartesian coordinate system through the points on the border between the empty and the full zone, are piecewise linear functions that do not decrease[9]. The question then is: "which minimum level of the determinant X is necessary for which level of outcome Y"? Figure 2 (top) shows that for an outcome (performance) of Y=2, it is necessary that the determinant (resource/effort) has value X=2. However, X=2 does not guarantee Y=2. Therefore, X=2 is necessary but not sufficient for Y=2. For an outcome of Y=1, it is necessary that the determinant has value of at least X=1. However, X≥1 does not guarantee Y=1. Therefore, X≥1 is necessary but not sufficient for Y=1.
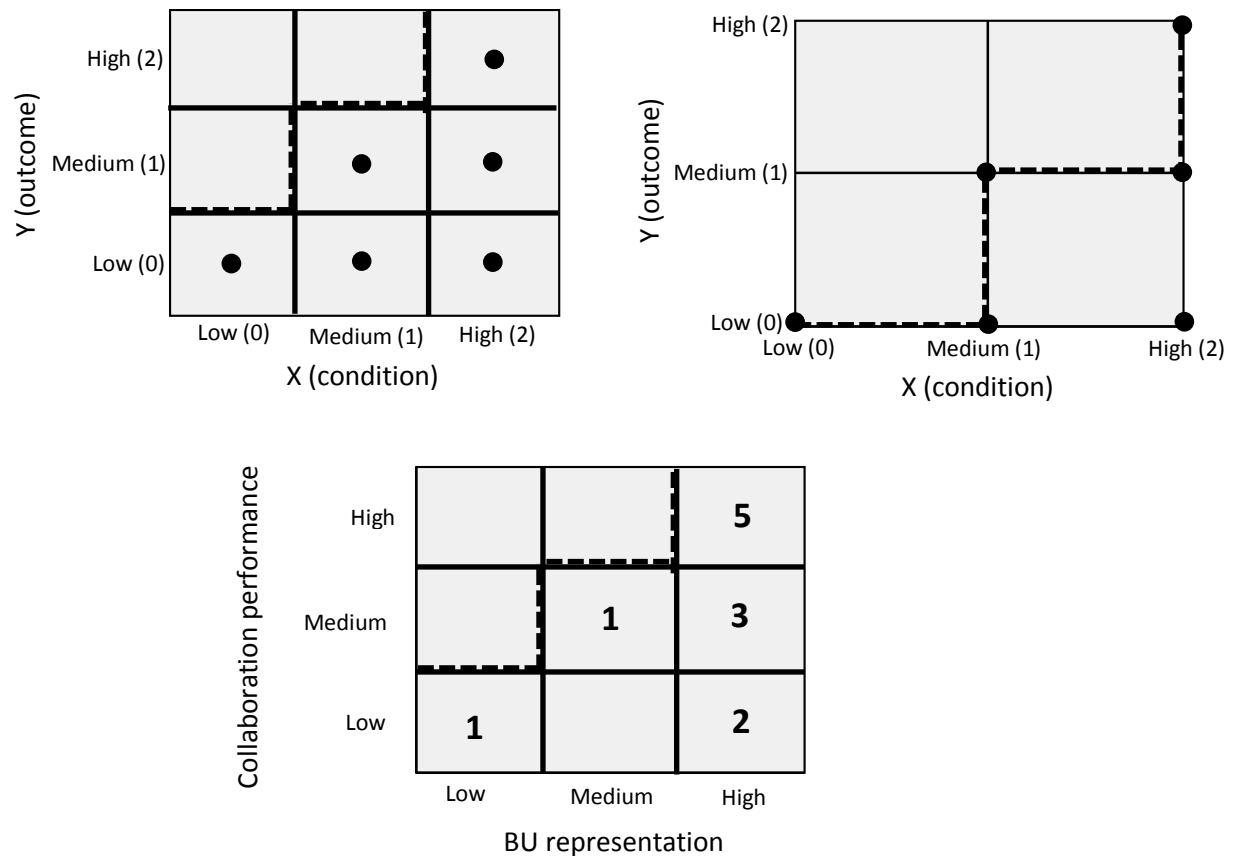
Figure 2. The discrete necessary condition with three levels Low (0), Medium (1), High (2) in a contingency matrix (top left) and in the grid of a Cartesian coordinate system (top right). The dashed line is the ceiling line. At the bottom is an example of a contingency table with X is Business unit representation in a collaboration team (Low, Medium or High) and Y Collaboration performance. (Data from Martin & Eisenhardt, 2010; see Appendix 2).

Figure 2 (top) shows that the value X=0 is a constraint for reaching both Y=1 and Y=2. The value X=1 is a constraint for reaching only Y=2. If the middle medium/medium cell (top left) or point (top right) were absent, the constraint of X=1 would be stronger, because X=1 would then also constrain Y=1. The size of the empty zone is an indication of the strength of the constraint of X on Y (more details follow in subsection Effect Size of a Necessary Condition). Also for testing or finding discrete necessary conditions, cases can be selected purposively on the bases of the presence of the dependent variable. If one is interested in necessary conditions for Y=2, then cases where this outcome is present (successful cases: Y=2) are selected and the level of the condition X is observed. If in all successful cases no X is smaller than X=2, X=2 can be considered as necessary for Y=2. Similarly if one is interested in necessary conditions for Y=1, then cases with this outcome are

selected. If in all these cases no X is smaller than X=1 (i.e., X≥1), then it can be considered that X≥1 is necessary for Y=1. The above analyses of discrete necessary conditions can be applied to both ordinal variables (with Figure 2, top left) and interval variables (with Figure 2, top left or right; the Cartesian coordinate system assumes known order and distances between the category values), and can be extended with more discrete categories. A discrete necessary condition with a large number of categories nears a continuous necessary condition.

Illustrative Example of the Discrete Necessary Condition

In a multiple case study, Martin and Eisenhardt (2010) analyzed success of collaboration between business units in multi-business firms. From their data a discrete (trichotomous) necessary condition on the relationship between "Business unit representation" in the collaboration team and "Collaboration performance" can be derived (see Appendix 2), which is shown in Figure 2 (bottom). It turns out that a "High" level of BU representation is necessary but not sufficient for a "High" level of collaboration performance. All five highly successful cases (collaborations) have this high level of BU representation. There are also five other cases with "High" level BU representation, but these cases apparently did not achieve the highest level of performance. If only a "Medium" level of collaboration performance were desired, then at least a "Medium" level of BU representation is necessary, but again this is not sufficient because two collaborations had higher than medium levels of BU representation, and showed "Low" performance. This analysis results in two *necessary (but not sufficient)* propositions: "A high level of BU representation is necessary but not sufficient for a high-performing cross-business unit collaboration", and "At least a medium level of BU representation is necessary but not sufficient for a medium level of cross-business unit collaboration performance." These propositions are different and more specific than the corresponding traditional proposition proposed by Martin and Eisenhardt (2010), which are implicitly formulated in terms of sufficiency: "BU representation *is more likely to yield* high collaboration performance." Other examples of necessary condition hypotheses derived from Martin and Eisenhardt's (2010) data set are presented in Appendix 2.

In comparison to the dichotomous necessary condition analysis, the discrete necessary condition adds more variable levels to the analysis and can therefore provide more specific propositions. The trichotomous necessary condition in Figure 2 shows which levels of the determinant (medium or high) are necessary for reaching the desired level of collaboration performance (medium or high). Sometimes a medium level of the determinant may be enough for allowing the highest level of performance. Then, a higher level of the determinant may be

"inefficient" (a further discussion on necessity inefficiency is presented in subsection Necessity Inefficiency).

Continuous Necessary Condition

In the case of a continuous necessary condition, the determinant X and the outcome Y are continuous variables (ratio variables). This means that the condition and the outcome can have any value within the limits of the lowest (0%) and highest (100%) values, allowing for even further detail[10]. Figure 3 contains graphs that correspond to the continuous necessary condition. The empty zone without observations is separated from the zone with observations by a straight ceiling line. In general, a ceiling is a boundary that splits a space in an upper part and a lower part, with only feasible points in the lower part of the space. In the two-dimensional Euclidian space (XY-plane), the expression "X is necessary but not sufficient for Y" is graphically represented by the ceiling line $Y = f(X)$, with inequality $Y \leq f(X)$ representing the feasible points in the lower part of the space (on and below the ceiling line). For the linear two-dimensional case the ceiling line is a straight line $Y = aX + b$ (a=slope; b=intercept) with $Y \leq aX + b$ representing the lower part with the feasible XY-points. In this paper, it is assumed that the ceiling line is linear and does not decrease[11].

The continuous necessary condition logic implies that a certain desired level of the outcome ($Y = Y_c$) can only be achieved if $X \geq X_c$. This means that X_c is necessary for Y_c, but X_c is not sufficient for Y_c because there may be observations $X = X_c$ with $Y < Y_c$, hence there are observations below the ceiling line for a given X_c. Within the scope of possible XY-values, the feasible "necessary but not sufficient" space in Figure 3 (top left) is triangular: the ceiling line coincides with the diagonal of the scope. Although it is frequently stated that the necessary but not sufficient condition can be graphically represented by a triangular shape of the feasible area (Goertz and Starr, 2003; Ragin, 2000) this shape is just a special case of the more common pentagonal shape (Figure 3, top right). In the present paper we focus on the general linear ceiling line with a typical pentagonal feasible space. In traditional approaches, heteroskedasticity in such scatter plots (increasing variance with increasing X) is considered a concern for performing standard regression analysis, whereas it is inherent to the necessary but not sufficient logic. Traditional approaches draw average lines through the middle of the data (average trends), whereas NCA draws borderlines between the zone with, and the zone without observations in the left upper hand corner. Therefore, traditional approaches to analyze pentagonal scatterplots may disconfirm "X causes Y" because the regression line may

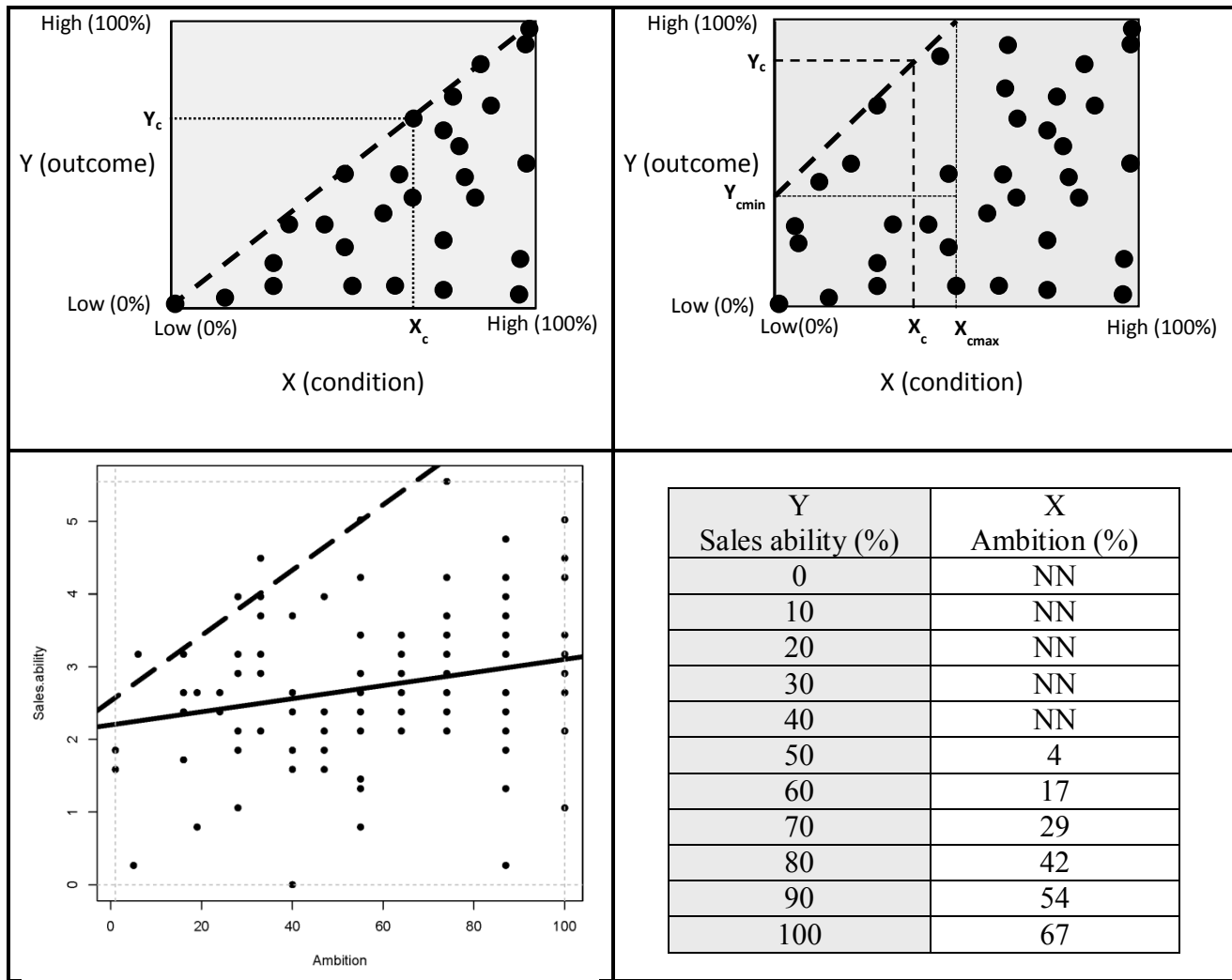| Y Sales ability (%)	X Ambition (%)
0	NN
10	NN
20	NN
30	NN
40	NN
50	4
60	17
70	29
80	42
90	54
100	67

Figure 3. The continuous necessary condition with a large number of X and Y levels between Low (0%) and High (100%), plotted in a Cartesian coordinate system. The dashed line is the ceiling line.

Top left: Idealized ("triangular") scatterplot: the ceiling line coincides with the diagonal.

Top right: Common ("pentagonal") scatterplot: the ceiling line intersects $X=X_{Low}$ and $Y=Y_{High}$ axes at other points than $[X_{Low},Y_{Low}]$ and $[X_{High},Y_{High}]$. The upper left area between the lines $X= X_{cmax}$ and $Y= Y_{cmin}$ is the *necessary condition zone* where each X constrains Y, and each Y is constrained by X.

Bottom left: Example[12] of a continuous necessary but not sufficient condition: Employee ambition is necessary but not sufficient for sales ability. The selected ceiling line technique (CR-FDH) allows some data points above the ceiling line (see subsection Ceiling Techniques). The solid line through the middle of the data is the OLS regression line.

Bottom right: Bottleneck table for the example illustrating what minimum level of the necessary conditions (X=Ambition) is required for different desired levels of the outcome (Y=Sales ability). NN=Not Necessary.

be virtually horizontal, but the NCA approach may confirm "X causes Y" because there is an empty zone in the upper left hand corner. The presence and size of an empty zone is an indication of the presence of a necessary condition. The empty zone in the upper left hand corner reflects the constraint that X puts on Y. For example, the value $X=X_c$ is a constraint (bottleneck) for reaching level of $Y>Y_c$. A higher level of Y is only possible by increasing X. Also for testing or inducing *continuous* necessary conditions, cases can be selected purposively on the bases of the presence of the dependent variable. If one is interested in necessary conditions for $Y>Y^*$, then cases where this outcome is present (successful cases: $Y>Y^*$) are selected, the cases are plotted in a Cartesian coordinate system, and a ceiling line is drawn to represent the necessary condition X_c for Y_c (see the subsection Ceiling Techniques).

Illustrative Example of the Continuous Necessary Condition

The Hogan Personality Inventory (HPI) is a widely used tool to assess employee personality for predicting organizational performance (Hogan & Hogan, 2007). Figure 3 (bottom left) shows an example of the relationship between the HPI ambition score (one's level of drive and achievement orientation), and supervisor rating of sales ability in a dataset of 108 sales representatives from a large food manufacturer in the U.S.A. Employees completed the HPI and supervisors rated employee's performance in term of sales ability using a seven-item reflective scale (e.g., "seeks out and develops new clients").

A traditional data analysis of the relationship between ambition and sales ability would consist of an OLS regression line through the middle of the data (solid line in Figure 3, bottom left). This line runs slightly upwards indicating a small positive average effect of ambition on sales ability ($R^2 = 0.07$). The data justify the conclusion that on average more ambition results in more sales ability, but this is possibly not practically relevant because of the small effect size. However, the necessary but not sufficient interpretation of the data is that the empty zone in the upper left hand corner indicates that there is a necessary condition. The ceiling line indicates that there is a ceiling on sales ability depending on the sales representative's level of ambition. The CR-FDH technique (see subsection Ceiling Techniques) that is used in this example to draw ceiling lines allows some data points above the ceiling line. For a desired sales ability level 4 it is necessary to have an ambition level of at least 30. About 20 percent of the sales representatives in the sample do not reach this level of ambition, which suggests that people with ambition levels below 30 will certainly fail to reach a sales ability level of 4 of higher. The other, about 80%, sales representatives who do

have an ambition level of at least 30 could reach such a high level of sales ability, but this is not guaranteed; in fact only about 10% of the sales representatives in the sample obtained this level of sales ability. In other words, an ambition level of at least 30 is necessary but not sufficient for a sales ability level of at least 4. An ambition level below 30 is a bottleneck for this high level of sales ability. Figure 3 (bottom right) shows the "Bottleneck technique", which is a table that indicates which level of the necessary condition (Ambition) is necessary for a given desired level of outcome (Sales ability in %), according to the ceiling line.

Multivariate Necessary Condition

Until now the analysis was bivariate with one X that is necessary for one Y. In this subsection, a more complex situation is considered: more than one necessary condition (X_1, X_2, X_3, etc.), and one Y. Multicausality is common in the organizational sciences. For example, Finney and Corbett (2007) reviewed empirical research on Enterprise Resource Planning (ERP) implementation success and identified 26 organizational determinants of success. Evanschitzky, Eisend, Calantone, and Jiang (2012) performed a meta-analysis of empirical research and identified 20 determinants of product innovation success. These examples can be extended with many other examples from virtually any subfield in the organizational sciences: a large set of determinants contribute to the outcome, and the underlying studies utilize traditional approaches such as correlation and regression to identify these determinants. Then, the size of the correlation or regression coefficient indicates the importance of a determinant. In this traditional logic, several determinants (X_1, X_2, $X_{3\dots}$) contribute to the outcome and can compensate for each other. In necessary but not sufficient logic, each single necessary determinant X_i always needs to have its minimum level X_{ic} to allow Y_c, independent of the value of the other determinants. If X_i drops below X_{ic}, the other determinants cannot compensate it. After the drop, the desired outcome Y_c can only be achieved again if X_i increases towards X_{ic}. This absence of a compensation mechanism once again shows the practical importance of identifying necessary conditions: putting minimum required levels of it in place, and keeping these levels in place: "Satisfying necessary conditions [...] constitutes the foundation of achieving the goal" (Dettmer, 1998, p.6).
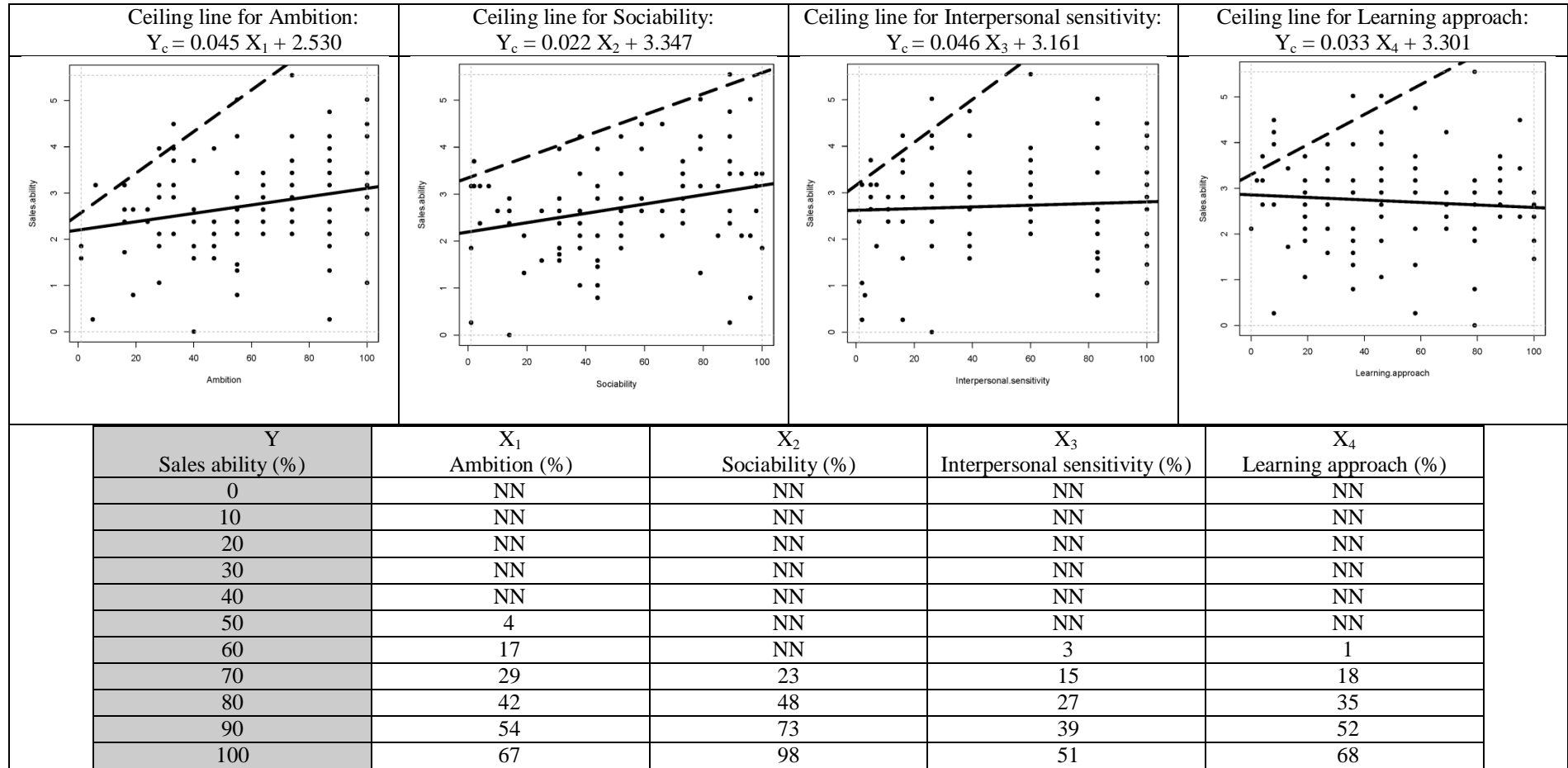
Illustrative Example of the Multivariate Necessary Condition

As an example, the upper part of Figure 4 shows the scatterplots of four personality scores measured with the Hogan Personality Inventory in the same sample as presented above. The plot for the first HPI score Ambition is the same as Figure 3 (bottom left). The

other three scatterplots are for HPI scores Sociability, Interpersonal sensitivity, and Learning approach. The regression lines (solid lines through the middle) show that all four personality traits have a relatively small effect on Sales ability. Ambition and Sociability have a small positive effect. The Interpersonal sensitivity regression line is nearly horizontal. The Learning approach regression line is slightly negative (!), indicating that on average Learning approach has a small negative effect on Sales ability. A traditional multiple regression analysis indicates that the four personality traits together hardly predict performance (adjusted $R^2 = 0.10$). Yet, all four plots, including the plot for Learning approach, show an empty space in the upper left hand corner, indicating that each personality trait may be necessary but not sufficient for Sales ability. The necessary condition analysis predicts failure in these zones: the level of Y cannot be reached with the corresponding level of X. This applies to each determinant separately.

The empty zones have different shapes because the ceiling lines (dashed lines drawn by using the CR-FDH ceiling technique, see subsection Ceiling Techniques) have different intercepts and slopes (see the ceiling equations in Figure 4, top). It implies that the necessary determinants put different constraints on the outcome[13]. This is illustrated with the multivariate bottleneck technique. The bottleneck table in the lower part of Figure 4 shows that for a desired level of outcome $Y_c = 80$ (80% of the maximum observed score), the necessary levels of X_1, X_2, X_3, X_4 are different: $X_{1c}=42$, $X_{2c}=48$, $X_{3c}=27$, and $X_{4c}=35$. A few important points are worth noting. First, we should recognize that each X variable must obtain these minimum ceiling values ($X_i \geq X_{ic}$) before it is possible to obtain $Y_c=80$. Thus, if even one of the X_i values falls below its ceiling value, $Y_c=80$ will not be reached. This contrasts with traditional applications of the general linear model, where the X variables are allowed to operate in a compensatory manner (e.g., a low score on X_1 may be compensated for with a higher score on X_2). Second, it is also important to recognize that the ceiling values will change if we change the level of the desired outcome. For $Y_c = 100$, $X_{1c}=67$, $X_{2c}=98$, $X_{3c}=51$, and $X_{3c}=68$. Finally, it is also possible that for a given (lower) desired value of the outcome, one or more determinants are no longer necessary. For example, for $Y_c = 60$ or lower, determinant X_2 (Sociability) is no longer a bottleneck. For $Y_c = 50$ or lower, also X_3 (Interpersonal sensitivity) and X_4 (Learning approach) are no longer a bottleneck, and none of the determinants are necessary for $Y_c \leq 40$.

It is typical for multivariate necessary conditions that no determinant is necessary at relatively low levels of the desired outcome, whereas several determinants become necessary at higher levels of desired outcome. Which specific determinant is the bottleneck depends on the specific levels of the determinants and the desired outcome. For example, if each of the

| | Ceiling line for Ambition: $Y_c = 0.045 X_1 + 2.530$ | Ceiling line for Sociability: $Y_c = 0.022 X_2 + 3.347$ | Ceiling line for Interpersonal sensitivity: $Y_c = 0.046 X_3 + 3.161$ | Ceiling line for Learning approach: $Y_c = 0.033 X_4 + 3.301$ |

Y Sales ability (%)	X_1 Ambition (%)	X_2 Sociability (%)	X_3 Interpersonal sensitivity (%)	X_4 Learning approach (%)
0	NN	NN	NN	NN
10	NN	NN	NN	NN
20	NN	NN	NN	NN
30	NN	NN	NN	NN
40	NN	NN	NN	NN
50	4	NN	NN	NN
60	17	NN	3	1
70	29	23	15	18
80	42	48	27	35
90	54	73	39	52
100	67	98	51	68

Figure 4. Example [12] of the multivariate necessary condition with four necessary conditions. The dashed lines are ceiling lines. The selected ceiling line technique (CR-FDH) allows some data points above the ceiling line (see subsection Ceiling Techniques). The solid line is the OLS regression line.

Upper part: Ceiling lines for four necessary conditions (Ambition, Sociability, Interpersonal sensitivity, Learning approach).

Lower part: Bottleneck table: required minimum levels of the necessary condition for different desired levels of the outcome (Sales ability). NN=Not Necessary.

four determinants has a score of 20, such an Ambition score and Sociability score allow a Sales ability of at least 60, whereas such an Interpersonal sensitivity score and Learning approach score allow a Sales ability of at least 70. Hence, a desired Sales ability of 60 is possible with scores of 20 of the four HPI determinants, but a desired Sales ability of 70 is not possible with such scores: the Ambition and Sociability scores are bottlenecks for such an outcome. The outcome of 70 is only possible if the Ambition score increases from 20 to at least 29, and the Sociability score from 20 to at least 23. The bottleneck table also shows that for a desired Sales ability score below 40 no HPI score is a bottleneck, which illustrates "outcome inefficiency" and indicates that the outcome could be higher, even with low HPI scores (for a discussion on outcome inefficiency see subsection Necessity Inefficiency) . The table also shows that for the highest outcome level (sales ability=100) condition inefficiency is present: all HPI scores are smaller than 100, with only the Sociability score close to 100. This indicates that for reaching the highest level of outcome, only about half to two-thirds of the maximum HPI scores for Ambition, Interpersonal sensitivity, and Learning approach are necessary, whereas for Sociability the maximum HPI score is necessary (for a discussion on condition inefficiency see subsection Necessity Inefficiency).

Multivariate necessary condition analysis with the bottleneck technique identifies which determinants, from a set of necessary determinants, successively become the weakest links (bottlenecks, constraints) if the desired outcome increases. In other words, for a given level of the desired outcome, multivariate necessary condition analysis identifies the necessary (but not sufficient) minimum values of the determinants to make the desired outcome possible.

Necessary Condition Analysis (NCA)

The continuous necessary condition case can be considered as the general case that includes the dichotomous and discrete cases (the graphs on the right hand sides in Figures 1 and 2 are essentially the same as the graphs in Figure 3). Therefore, the general methodology Necessary Condition Analysis (NCA) is developed for determining necessary (but not sufficient) conditions. This methodology consists of two main parts: (1) determining ceiling lines and the corresponding bottleneck tables, and (2) calculating several parameters such as "accuracy" of the ceiling line, "effect size" of the necessary condition, and "necessity inefficiency". First, I propose and evaluate two classes of ceiling techniques for drawing ceiling lines: ceiling envelopment and ceiling regression, compare these techniques and some

other techniques, and present an example. Next, I develop an effect size measure for necessary conditions, and propose benchmark values, discuss "necessity inefficiency", and compare NCA with QCA. Finally, I discuss the limitations of NCA.

Ceiling Techniques

The starting point for the necessary condition analysis is a scatter plot of data using a Cartesian coordinate system, which plots X (the determinant and potential necessary condition) against Y (the outcome) for each case. If visual inspection suggests the presence of an empty zone in the upper left hand corner (with the convention that the X-axis is "horizontal" and the Y-axis is "vertical" and that values increase "upwards" and "to the right"), a necessary condition of X for Y may exist. Then a ceiling line between the empty zone without observations and the full zone with observations can be drawn. A ceiling line must separate the empty zone from the full zone as accurately as possible. Accuracy means that the empty zone is as large as possible without observations in it. However, drawing the best ceiling line usually requires making a trade-off decision between the size of the empty zone and the number of observations in the empty zone ("exceptions", "outliers", "counterexamples"). Because the "empty" zone may not always be empty, it is named "ceiling zone" (C).The position of the data points around the ceiling line might suggest that the best ceiling line is not linear, nor increasing. Additionally, the best ceiling line may be a smooth line or a piecewise (linear) function. A ceiling function can be expressed in general terms as $Y_c = f(X_c)$. For simplicity, the focus in this paper is on continuous or piecewise linear, non-decreasing ceiling lines, which appears acceptable for first estimations of ceiling lines for most datasets. A non-decreasing (constant or increasing) ceiling line is typical and allows for a formulation that $X \geq X_c$ is necessary for Y_c.

Goertz et al. (2013) made the first attempts at drawing ceiling lines. They visually inspected the ceiling zone of a specific scatterplot and divided it manually into several rectangular zones. Rather than defining the ceiling line, the combination of rectangles resulted in a piecewise linear ceiling function with all observations on or below the ceiling. The advantage of this technique is that it implicitly only uses the location of upper left observations (where the border exists) to draw the ceiling line and that it can be made accurate with all observations on or below the ceiling line. One disadvantage of this technique is that it is informal and based on subjective selection of rectangles. Additionally, Goertz et al. (2013) introduced quantile regression to systematically draw continuous linear ceiling lines. Quantile regression (Koenker, 2005) uses all observations to draw lines. The 50th quantile

regression line splits the observation in about half above and half below the line. This line is similar to a regular OLS regression line. However, for quantiles above, for example, 90, quantile regression results in a ceiling line that allows some points above the line. The advantage of this technique is that it is an objective, formal methodology to draw ceiling lines. However, a disadvantage is that it uses all observations below the ceiling line. For drawing a line between zones with and without observations, there is no rationale for using observations far away from this line.

For NCA the advantages of the two approaches are combined by proposing two alternative classes of techniques, Ceiling Envelopment (CE), and Ceiling Regression (CR). Both are formal techniques and use only observations close to the ceiling zone. CE is a piecewise linear line and CR a continuous linear line (straight line).

Ceiling envelopment. In this paper, Ceiling Envelopment is created for NCA on the basis of Data Envelopment Analysis (DEA) techniques (Charnes, Cooper, & Rhodes, 1978). DEA is used in operations research and econometrics to evaluate production efficiency of decision making units. Here the method is used to draw ceiling lines. Ceiling Envelopment pulls an envelope along upper left observations using linear programming. Only a few assumptions need to be made regarding the shape of the ceiling line, depending on the specific envelopment technique. The ceiling envelopment technique with Varying Return to Scale (CE-VRS) assumes that the ceiling line is convex. It results in a piecewise-linear convex ceiling line. The envelopment technique with Free Disposal Hull (CE-FDH) is a more flexible technique (Tulkens, 1993). It does not require any assumption regarding the ceiling line. Because of its flexibility, and intuitive and simple applicability to dichotomous, discrete, and continuous necessary conditions, CE-FDH is proposed as a default ceiling envelopment technique for NCA. It results in a piecewise linear function along the upper left observations. Drawing the CE-FDH ceiling line can be formulated in words as follows:

1. Start at point $Y=Y_{min}$ for observation $X=X_{min}$.

2. Move vertically upwards to the observation with the largest Y for $X=X_{min}$ (there can be more than one observation with the same X value, particularly for discrete variables).

3. Move horizontally to the right until the point with an observation on or above this horizontal line (discard observations below this line).

4. Move vertically upwards to the observation with the largest Y for this X (there can be more than one observations with the same X value, particularly for discrete variables).

5. Repeat 3 and 4 until the horizontal line in 3 has reached the $X=X_{max}$ line.

Ceiling regression. In this paper, Ceiling Regression (CR) is created for NCA on the basis of the outcomes of a CE analysis. Therefore two versions exist: one based on the outcomes of CE-VRS, and one based on the outcomes of CE-FDH. Ceiling regression CR-VRS draws an OLS regression line through the points that connect the linear pieces of the CE-VRS ceiling line, whereas CR-FDH draws an OLS regression line through the upper-left edges of the CE-FDH piecewise linear function i.e., the points where a vertical part of the CE-FDH line ends and continues as a horizontal line when X increases. These approaches smoothen the piecewise linear ceiling lines from the ceiling envelopment techniques by using a straight line. The straight line allows for further modeling and analyses, and for comparison of ceiling lines. For the same reasons that CE-FDH is preferred over CE-VRS, CR-FDH is preferred over CR-VRS.

Comparison of ceiling techniques. Table 2 compares the four different ceiling techniques. It also includes some other common techniques in operations research and econometrics for the estimation of efficiency frontiers that could be used for drawing ceiling lines as well. Corrected Ordinary Least Squares (COLS) and Stochastic Frontier Analysis (SFA) (e.g., Aigner & Chu, 1968; Bogetoft & Otto, 2011) have some disadvantages compared to the techniques selected above. COLS shifts the OLS regression line parallel towards the most upper point, and therefore uses all observations. SFA focuses on the observations around the ceiling line, but includes a stochastic term based on a probability assumption of the observations below the ceiling, and thus also uses all observations. A last technique called Low-High (LH) was created for the purpose of NCA as a rough first estimation of the ceiling line. It is the diagonal of the necessary condition zone defined by two specific observations in the dataset: the observation with the lowest X (with highest Y) and the observation with the highest Y (with lowest X). However, this technique is sensitive to outliers and measurement error.

The table shows that only five techniques use upper left points (LH, CE-VRS, CE-FDH, CR-VRS, and CR-FDH) to draw the border line between zones with and zones without observations. Within this group, LH uses only two observations, whereas CE-VRS, CE-FDH, CR-VRS and CR-FDH use observations close to the border line. When comparing CE-VRS and CE-FDH, the latter has fewer limitations because it does not require convexity of the ceiling function, and is therefore preferred. Because CR-VRS and CR-FDH are based on CE-VRS and CE-FDH respectively, CR-FDH also has fewer limitations. Therefore, CE-FDH is the default as a non-parametric technique (no assumption about ceiling line function), and CR-FDH is the default for a parametric technique (linear ceiling line). When comparing

Table 2

Comparison of Techniques for Drawing Ceilings Lines.

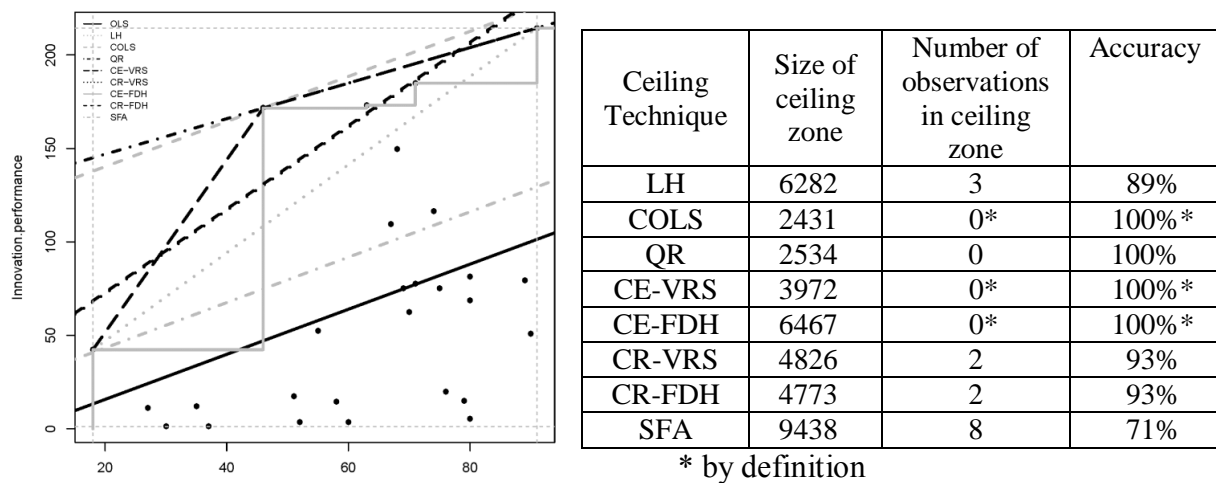| Ceiling line technique[1]	Observations that are used	Pre-defined ceiling function	Allows points above ceiling	Drawing procedure
CE-VRS	Upper left	No[2]	No	Optimization
CE-FDH	Upper left	No	No	Optimization
CR-VRS	Upper left	Yes	Yes	Optimization+Statistical
CR-FDH	Upper left	Yes	Yes	Optimization+Statistical
COLS	All	Yes	No	Statistical
QR	All	Yes	Yes	Statistical
SFA	All	Yes	Yes	Statistical
LH	Lowest left and highest right	Yes	Yes	Mathematical

[1] LH = Low-High; COLS = Corrected Ordinary Least Squares; QR = Quantile Regression; CE-VRS = Ceiling Envelopment – Varying Return to Scale; CE-FDH = Ceiling Envelopment – Free Disposal Hull; CR-VRS = Ceiling Regression – Varying Return to Scale; CR-FDH = Ceiling Regression – Free Disposal Hull; SFA= Stochastic Frontier Analysis.
[2] Convexity required

CE-FDH and CR-FDH, the former is preferred when a straight ceiling line does not properly represent the data along the border between the empty and full zone and when smoothing considerably reduces the size of the ceiling zone. This holds particularly for dichotomous variables, and for discrete variables with a small number of variable levels (e.g., max 5; see Figure 2). This may also happen if the number of observations is relatively low, i.e., in small datasets. If a straight ceiling line is an acceptable approximation of the data along the border, the CR-FDH technique is preferred. Compared to CE-FDH, CR-FDH usually has some observations in the ceiling zone due to the smoothing approach, and the ceiling zone may be somewhat smaller. By definition, CE-FDH does not have observations above the ceiling line; it is therefore more sensitive to outliers and measurement errors than CR-FDH.

Accuracy. The accuracy of a ceiling line is defined as the number of observations that are on or below the ceiling zone divided by the total number of observations, multiplied by 100%. Then, by definition, the accuracy for CE-VRS, CE-FDH and COLS is 100%, and for the other techniques the accuracy can be below 100%.

Example. In Figure 5, the eight techniques are applied to a dataset of 28 countries with data about a country's level of individualism according to Hofstede (1980), and a country's innovation performance according to the Global Innovation Index (Gans & Stern, 2003).

Ceiling Technique	Size of ceiling zone	Number of observations in ceiling zone	Accuracy
LH	6282	3	89%
COLS	2431	0*	100%*
QR	2534	0	100%
CE-VRS	3972	0*	100%*
CE-FDH	6467	0*	100%*
CR-VRS	4826	2	93%
CR-FDH	4773	2	93%
SFA	9438	8	71%

* by definition

Figure 5. Application of different ceiling techniques to a dataset on a country's individualism (X-axis) and its innovation performance (Y-axis). LH= Low-High; COLS=Corrected Ordinary Least Squares; QR=Quantile Regression; CE-VRS= Ceiling Envelopment – Varying Return to Scale; CE-FDH= Ceiling Envelopment – Free Disposal Hull; CR-VRS = Ceiling Regression – Varying Return to Scale; CR-FDH = Ceiling Regression – Free Disposal Hull; SFA= Stochastic Frontier Analysis The lowest solid line is the OLS regression line.

The scatterplot shows that there is an empty space in the upper left hand corner indicating the presence of a necessary condition: Individualism (X) is necessary for innovation performance (Y). The ceiling lines differ considerably across techniques. The techniques that use all 28 observations (i.e., COLS, QR, and SFA) seem to underestimate (COLS and QR) or overestimate (SFA) the size of the ceiling zone. From the remaining techniques that do not allow observations in the ceiling zone (CE-VRS and CE-FDH), CE-FDH defines a considerably larger ceiling zone than CE-VRS because it is not restricted by a convexity requirement. From the remaining techniques that allow observations in the ceiling zone and have a predefined (linear) ceiling line (LH, CR-VRS, CR-FDH), LH is less accurate than CR-VRS and CR-FDH, but defines a larger ceiling zone. In this example CR-VRS and CR-FDH result in virtually the same ceiling line. For reference the OLS line is shown in Figure 5 as well (the lowest straight line). It is clear that a line through the middle of the data that describes an average trend is not a proper candidate for a ceiling line (accuracy in this example is only 64%).

Analyzing a number of empirical samples and using simulated datasets indicate that CE-FDH and CR-FDH generally produce stable results with relatively large ceiling zones and with no (CE-FDH) or few (CR-FDH) observations in the ceiling zone. CE-FDH is similar to Goertz et al.'s (2013) rectangular approach and produces a piecewise linear function with no observations in the ceiling zone; CR-FDH is a new technique developed for the present purpose of NCA and is a combination of data envelopment and regression analysis. It produces a smooth linear ceiling function with normally a few observations in the ceiling zone. The ceiling techniques presented here are only linear non-decreasing ceiling lines. In future work also non-linearities will be explored e.g., triangular or trapezoidal piecewise linear functions, or non-linear functions such as polynomials or power functions. The quality of ceiling lines need to be systematically compared by evaluating fit, effect size, accuracy, and required assumptions.

Effect Size of the Necessary Condition

Perhaps in nearly all scatterplots, an empty space can be found in the upper left hand corner, which may indicate the existence of a necessary condition. The question then is: is the necessary condition large enough to be taken seriously? Therefore, there is a need to formulate an effect size measure. An effect size is a "quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest" (Kelley & Preacher, 2012, p.140).Applied to necessary conditions, the effect size should represent how much a given value of the necessary condition X_c constrains Y. The effect size measure for necessity but not sufficiency proposed here builds on the suggestion by Goertz et al. (2013, p.4) that "the importance of the ceiling …is the relative size of the no observation zone created by … the ceiling." The effect size of a necessary condition can be expressed in terms of the size of the constraint that the ceiling poses on the outcome. The effect (constraint) is stronger if the ceiling zone is larger. Hence, the effect size of a necessary condition can be represented by the size of the ceiling zone compared to the size of the entire area that can have observations. This potential area with observations is called the "scope" (S). The larger the ceiling zone compared to the scope, the lower the ceiling, the larger the ceiling effect, the larger the constraint, and therefore the larger the effect size of the necessary condition. The effect size can be expressed as follows: d=C/S, where d is the effect size[14], C is the size of the ceiling zone, and S is the scope. Hence, d is the proportion of the scope above the ceiling. The scope can be determined in two different ways: theoretically, based on theoretically expected

minimum and maximum values of X and Y, and empirically, based on observed minimum and maximum values of X and Y. Thus: $S=(X_{max}-X_{min}) \times (Y_{max}-Y_{min})$.

Similar to other effect size measures such as the correlation coefficient r, or R^2 in regression, this necessary condition effect size ranges from 0 to 1 ($0 \leq d \leq 1$). An effect size can be valued as important or not, depending on the context. A given effect size can be small in one context and large in another. General qualifications for the size of an effect as "small", "medium", or "large" (e.g., Cohen, 1988) are therefore disputable. If, nevertheless, a researcher wishes to have a general benchmark for necessary condition effect size, I would offer $0 < d < 0.1$ as a "small effect", $0.1 \leq d < 0.3$ as a "medium effect", $0.3 \leq d < 0.5$ as a "large effect", and $d \geq 0.5$ as a "very large effect"[15]. Two recent papers with applications of necessary condition analysis have effect sizes between 0.1 and 0.3. One of these is in political science (Goertz et al., 2013), the other in business research (Van der Valk, Sumo, Dul, & Schroeder, 2015). In both papers, these effect sizes were considered as theoretically and practically meaningful. If a dichotomous classification is desired (e.g., for testing whether or not a necessary condition hypothesis should be rejected, or for making an "in kind" qualification of the absence or presence of a necessary condition), I suggest to use effect size 0.1 as the threshold. A distribution of more than 150 continuous necessary conditions effect sizes from different studies indicates that roughly 50% of the effect sizes are small ($d < 0.1$), about 40% are medium ($0.1 \leq d < 0.3$), and about 10% are large ($0.3 \leq d < 0.5$). Very large effect sizes ($d \geq 0.5$) can only be expected when the ceiling line is not a straight line (see the example in Figure 6). Hence, an "in kind" necessary condition hypothesis in the continuous case ("X is necessary for Y") is rejected if the effect size d is less than 0.1.

The effect size of the dichotomous necessary condition, as shown in Figure 1, can be either 0 or 1. Without observations in the upper left hand cell, X is necessary for Y, and the effect size is 1. If there were observations in the upper left hand cell, X would not be necessary for Y, and the effect size would be 0. However, if there are only a few observations in the upper left hand cell in comparison to the total number of observations two interpretations are possible. In the deterministic interpretation X is still not necessary for Y. In the more flexible stochastic interpretation, which I adopt in this paper, after evaluation of the outliers in the "empty" cell, X could be considered as necessary. Outliers can be a "substitutes" as in the example of the GRE score in Figure 1 (bottom), or can simply represent error. With a few outliers and depending on the context, one may decide to denote the condition as a "virtually" necessary. No general rule exists about the acceptable number of cases in the "empty" upper left cell. Dul et al. (2010) suggest (arbitrarily) that a maximum

number of 1 out of 20 observations (5%) is allowed in the "empty" cell. Hence, this decision rule is not based on effect size but on accuracy: Below 95% accuracy the "in kind" necessary condition hypothesis in the dichotomous case is rejected[16].

In the dichotomous case, no other effect size values than 0 or 1 are possible. Yet, the left hand side of Figure 1 suggests that the ceiling zone equals 1 (one empty cell) and that the scope equals 4 (four cells with possible observations), hence that the effect size is ¼. This is incorrect. The actual effect size of the dichotomous case can be observed more easily from the right hand side of Figure 1. This figure shows that the scope equals 1 x 1 = 1. Also, the area of the empty space equals 1 x 1 = 1, and therefore the effect size is 1. Similarly, in the discrete case of Figure 2 without medium-medium observations, both the empty space and the scope would be (2-0) x (2-0) = 4, thus the effect size equals 1. With medium-medium observations, the scope would be (2-0) x (2-0) = 4, and the ceiling zone is (1-0) x (2-0) + (2-1) x (2-1) = 3, hence, the effect size is ¾ (and not 3/9 as suggested by the left hand side of Figure 2)[17].

In the example shown in Figure 5, the minimum and maximum observed values of X (18, 91) and Y (1.2, 214.4) result in an empirical scope of 15563.6. Consequently, the effect size calculated with CE-FDH is 0.42, and the effect size calculated with CR-FDH is 0.31.

Necessity Inefficiency

The effect size is a general measure that indicates to which extend the "constrainer" X constrains Y, and the "constrainee" Y is constrained by X. However, normally, not for all values of X, X constrains Y, and not for all values of Y, Y is constrained by X. When the feasible space is triangular as in Figure 3 (top left) X always constrains Y, and Y is always constrained by X. But when the feasible space is pentagonal as in Figure 3 (top right) the ceiling line intersects the $Y=Y_{High}$-axis and the $X=X_{Low}$-axis at other points than $[X_{Low},Y_{Low}]$ and $[X_{High},Y_{High}]$. As a result, for $X>X_{cmax}$, X does not constrain Y, and for $Y<Y_{cmin}$, Y is not constrained by X. The upper left area between the lines $X=X_{cmax}$ and $Y=Y_{cmin}$ in Figure 3 (top right) is the "necessary condition zone" where X always constrains Y, and Y is always constrained by X. Within this zone, it makes sense to increase the level of the necessary condition in order to allow for higher levels of the outcome. Outside this zone, increasing the necessary condition level does not have such effect. The area outside the necessary condition zone can be considered as "inefficient" for necessity. Necessity inefficiency has two components. Condition inefficiency (denoted here as i_x) specifies that a level of resource/effort $X > X_{cmax}$ is not needed even for the highest level of performance Y[18]. Condition inefficiency can be calculated as $i_x = (X_{High} - X_{cmax})/(X_{High} - X_{Low})*100\%$, and can

have values between 0 and 100%. Outcome inefficiency (i_y) indicates that for a level of outcome $Y < Y_{cmin}$ any level of X allows for a higher value of the outcome (e.g., higher performance ambition). Outcome inefficiency can be calculated as $i_y = (Y_{cmin} - Y_{Low})/(Y_{High} - Y_{Low})*100\%$, with values between 0 and 100%. The effect size is related to inefficiency as follows: $d = \frac{1}{2}(1- i_x/100\%)*(1- i_y/100\%)$; The larger the inefficiencies, the smaller the effect size. When both inefficiency components are absent, the idealized situation of Figure 3 (top left) is obtained, and the effect size is 0.5, which is the theoretical maximum for a straight ceiling line. When one of the inefficiencies equals 100%, there is no ceiling, and thus no necessary condition and zero effect size. When the X and Y axes are standardized (e.g., X_{Low} and Y_{Low} are 0 or 0% and X_{High} and Y_{High} are 1 or 100%), the angle of the ceiling line is 45 degrees if both inefficiencies are equal with standardized values between 0 and 100%. When condition inefficiency is larger than outcome inefficiency, the ceiling line is steep (>45 degrees). A slope that is less steep (<45 degrees) indicates that outcome inefficiency is larger than condition inefficiency. In Figure 3 (bottom left), the condition inefficiency is 33%. This means that for an Ambition level (X) above 67% of 100 = 67, ambition is not necessary for even the highest level of Sales ability. In Figure 3 (bottom left) outcome efficiency is 46%. This means that for a desired Sales ability level (Y) that is below 46% of 5.5 (the maximum observed level) = 2.5, Ambition is not necessary for Sales ability. Because outcome inefficiency (46%) is larger than condition inefficiency (33%), the slope of the ceiling line is less than 45 degrees.[19]

Comparison NCA with QCA

Multicausal phenomena can also be analyzed with other approaches that use the logic of necessity and sufficiency. For example, Qualitative Comparative Analysis (QCA, Ragin, 2000, 2008; for an introduction see Schneider & Rohlfing, 2013) is gaining growing attention in organizational sciences (Rihoux, Alamos, Bol, Marx, & Rezsohazy, 2013). It aims to find combinations of determinants that are sufficient for the outcome, while separate determinants may not be necessary or sufficient for the outcome (Fiss, 2007). NCA and QCA are similar in the sense that they both approach causality in terms of necessity and sufficiency, and not in terms of correlation or regression. However, two major differences are that NCA focuses on (levels of) single determinants (and their combinations), whereas QCA focuses on combinations of determinants (configurations), and that NCA focuses on necessary determinants that are not automatically sufficient, whereas QCA focuses on sufficient configurations that are not automatically necessary (equifinality with several possible causal

paths)[20]. Although QCA centers on sufficient configurations it is also used for necessary condition analysis (e.g., Ragin, 2003). But the QCA approach for "continuous" necessary condition analysis (fuzzy set QCA) is fundamentally different than the NCA approach presented here (Vis and Dul, 2015). Instead of a ceiling line, QCA uses the bisectional diagonal through the theoretical scope as the reference line for evaluating the presence of a necessary condition. Thus QCA presumes that necessary conditions are only present in triangular scatter plots[21]. If all observations are on and below the diagonal, the QCA reference line is the same as the NCA ceiling line. However, when a substantial number of data points show up above the reference line (e.g., because of inefficiencies, see examples in Figure 3 (top right and bottom), and in Figure 4), NCA "moves the ceiling line upwards", possibly with rotation, while the QCA reference line remains the same (the diagonal). Then NCA concludes that X is necessary for Y at lower levels of X, whereas QCA concludes that X is "less consistent" with necessity (necessity "consistency", Bol & Luppi, 2013; Ragin, 2006) without specifying values of X where X is necessary for Y or where X is not necessary for Y. As a consequence, QCA normally finds considerably less necessary conditions in a datasets than NCA. Another difference is that QCA only analyses *in kind* necessary conditions ("condition X is necessary for outcome Y"), whereas NCA can also analyze *in degree* necessary conditions ("level of X is necessary for level of Y"). NCA and QCA are also different with respect to evaluating the "importance" of a necessary condition. In NCA importance is defined as the size of the empty zone in comparison to the scope, whereas in QCA (and other analytic approaches, e.g., Goertz, 2006b[22]) importance of a necessary condition is defined as the extent to which the necessary condition is also sufficient (necessity "coverage"). In this view, a necessary condition is more important if it is more sufficient[23]. However, a fundamental notion in the current paper is that single organizational ingredients are not sufficient for multicausal organizational phenomena, which explains and justifies the "necessary but not sufficient" expression. For a further discussion on the differences between NCA and QCA, see Vis and Dul (2015), and Dul (2015).

Applying NCA instead of QCA for analyzing necessity can result is different conclusions. For example, Young and Poon (2013) applied QCA to identify the necessity of five IT project success factors: top management support, user involvement, project methodologies, high level planning, and high quality project staff. Using a sufficiency view (Goertz, 2006b) of importance of necessity, they found importance levels of 0.97, 0.64, 0.70, 0.69, and 0.61, respectively. Applying the NCA approach to Young and Poon's dataset using CE-FDH gives necessary condition effect sizes of 0.30, 0.25, 0.19, 0.33 and 0, respectively.

Comparing the rank orders of these two approaches shows that top management support is the most important in the sufficiency view, followed by project methodologies and high level planning. However, in the necessary but not sufficient view, high level planning has the largest effect size, followed by top management support and user involvement. The bottleneck analysis indicates that only project methodologies is necessary for a medium success level of 0.5 (the observed success levels ranged from 0.1 to 0.9), and the other four determinants are not. Yet, if the desired level of success increases, more conditions become necessary, and four out of five conditions are necessary for the maximum observed outcome level (0.9), with required levels of the five determinants of at least 88%, 67%, 25%, 100%, and 0% of the maximum range of observed levels of these determinants, respectively (the last determinant is not necessary). These percentages show necessity inefficiencies for all four determinants that are necessary, except for high level planning that needs a level of 100% for maximum outcome. This NCA analysis adds to Young and Poon's (p. 953) insights that top management support "is significantly more important for project success than factors emphasized in traditional practice." The NCA analysis shows that top management support is needed if maximum performance is desired, but not at maximum level. Top management support is *not* needed if only medium performance is desired (then "project methodologies" is needed). More importantly, for maximum performance, adequate levels of the other determinants, which Young and Poon call the "factors emphasized in traditional practice" are also needed. Hence, high outcomes will not occur without properly balancing the minimum levels of the four necessary determinants. Therefore, NCA analysis does not support Young and Poon's (2013, p. 953) conclusion that "current practice emphasizing project methodologies may be misdirecting effort", because considerable levels of four necessary determinants including project methodologies are needed for high performance. This example of an additional NCA analysis is presented for purposes of illustration and it is typical for the necessary but not sufficient approach. It may provide new insights about how organizations can efficiently put and keep organizational ingredients in place to achieve a desired outcome, and by which level.

Limitations of NCA

Just like any other data analysis approach, NCA has several limitations. One fundamental limitation that it shares with other data analysis techniques is that NCA cannot solve the problem of "observational data cannot guarantee causality". Observational studies are widespread in organizational research. All examples in Table 1 and all the examples used

for illustration in this paper are from observational studies. Observing a data pattern that is consistent with the causal hypothesis is not evidence of a causal connection. Hence, it is important that identified necessary conditions are "theoretically justified" i.e., that it is understood how X constrains Y, and Y is constrained by X. Requirements for causal inference in empirical studies for building or testing necessary cause-effect relations are the same as for any other type of cause-effect relation. For example, a necessary cause is more plausible if the cause precedes the outcome, and is related to the outcome, and if an observed outcome cannot be explained by another cause (Shadish et al., 2002). If such requirements are not met, a necessary condition outcome that is empirically observed may, for example, be spurious, i.e., caused by an unobserved common variable (Spector & Brannick, 2010). This can be illustrated with a revision of the classic example of the relation between the number of storks in a region, and the number of human newborns in that region (Box, Hunter, & Hunter, 1978)[24]. Figure 6 shows a scatterplot of the relationship between the annual number of white stork nesting pairs and the number of human births in the Oldenburg region in Germany over one decade. The scatterplot shows that there is an empty area in the upper left hand corner. The ceiling line is drawn with the CE-FDH, rather than with the CR-FDH ceiling technique because the observations along the border between the empty zone and the full zone cannot be well represented by a straight line.

The data suggest the existence of a necessary condition with a very large effect size (0.73): a high number of human births in the region is only possible if the number of stork nesting pairs in the region is large. It is clear (unless one believes in fables or in the Theory of the Stork, Höfer & Przyrembel, 2004) that a theoretical explanation is missing in this example. A necessary relationship that is observed between X and Y can be spurious if a non-observed variable Z is sufficient for X, and necessary for Y (Mahoney, 2007). In the storks example one could state that when more people (Z) settle in a region and build houses with chimneys and other high human-made constructions, the number of stork nesting pairs (X) increases because storks favor high nesting places: Z produces X, hence is a sufficient cause of X. Additionally, Z is a necessary cause of Y because a high increase of the number of births is only possible when there are more people. In other words, the non-observed confounding variable Z is the reason of the presumed "necessary" relation between storks and births[25]. This example shows again that without theoretical support, a necessary condition cannot be presumed from observational data. This is not different from any other data analysis approach using observational data.
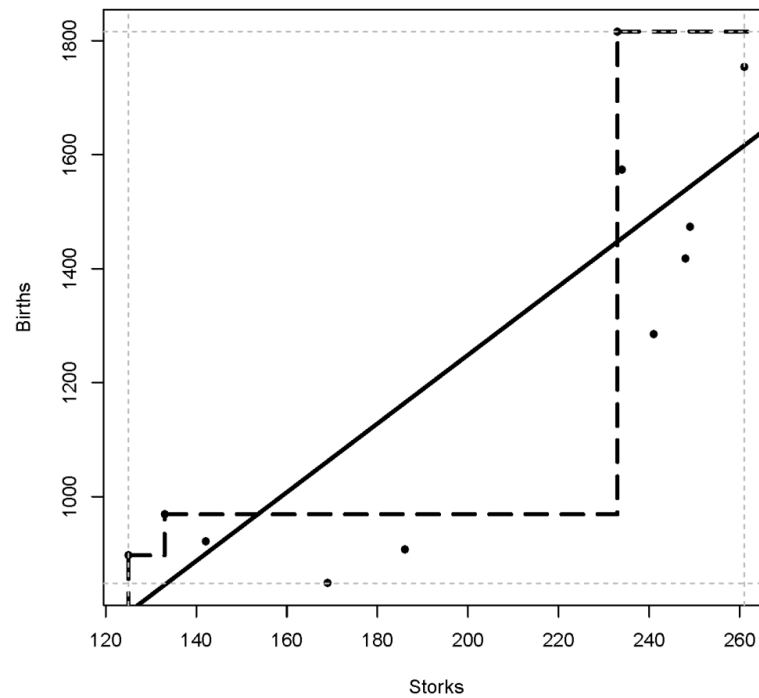
Figure 6. Scatterplot of the relation between the number of white storks nesting pairs in a region ("Storks") during a given year and the number of human births during that year ("Births"). The dashed line is the CE-FHD ceiling line, and the solid line is the OLS regression line. (After: Van Maanen, 2010).

If the cause can be manipulated, a research design based on randomized experiments may be better equipped to fulfill requirements for causality than an observational study. A potential necessary cause of a multicausal phenomenon (e.g., storks for birth) can be tested in a randomized experiment where instances (cases) *with* the cause (regions with storks) are selected randomly from a defined population (all regions with storks), and then randomly assigned to the experimental group (regions that get a treatment) or control group (regions that do not get a treatment). In the experimental group the presumed necessary causal condition is manipulated by *taking it away* (e.g., storks are chased away), whereas in the control group the presumed causal condition is unchanged. Then we observe whether the effect *disappears* in the experimental group (reduction of birth rate), but not in the control group (birth rate constant). Such a design provides strong evidence for the (non-)existence of a necessary cause. This type of experiment contrasts the common "sufficiency" experiment in which first cases *without* the cause (regions without storks) are randomly selected from a

population (all regions without storks) and randomly assigned to the experimental or control group. Then the cause is *added* to the experimental group (storks are introduced), and we observe if the effect *appears* in the experimental group (increase of birth rate) and not in the control group (birth rate constant) [26]. The necessity experiment can be rephrased as a sufficiency experiment because a necessary condition that is formulated as "the presence of the cause is necessary for the presence the effect" can also be formulated as "the absence of the cause is sufficient for the absence of the effect". Then in the sufficiency approach for testing a necessary cause, the "absence of the cause" (no storks) is added to the experimental group and we observe if the "absence of the effect" (reduction of births) appears. If the necessary cause cannot be manipulated (cannot be taken away), non-experimental research designs such as natural experiments (where the necessary cause disappears naturally, and we observe if the effect disappears) or observational studies (where we observe if the necessary cause is present in cases with the effect, or if the effect is absent in cases without the necessary cause) can be applied. Such approaches rely more on solid theoretical reasoning in order to infer necessity causality.

A limitation of NCA is that it may be more susceptible for sampling and measurement error than traditional data-analysis approaches. The reason is that the proposed ceiling techniques (CE-FDH and CR-FDH) only use a small proportion of the observations in the sample for drawing the ceiling line (the observations near the ceiling line). The ceiling line (and therefore other quantities derived from it (e.g., effect size, accuracy) may be sensitive to selection bias and measurement error in the cases around the ceiling line. This particularly holds if the number of cases around the ceiling line is relatively small. Therefore the accuracy of the ceiling line depends on having representative cases around the ceiling line with accurate measurement of the variables. Measurement precision and validity need to be ensured, in particular for the cases that are expected to be around the ceiling line, (e.g., relatively successful cases). The ceiling line may also be susceptible to exceptions, outliers, or counterexamples in the upper left hand corner of the scatterplot. Therefore, observations near the ceiling line need to be evaluated before conclusions can be drawn. Exceptions (high outcome without the necessary condition) may be due to measurement error, unexplained stochasticity in the data, the case not being part of the theoretical domain where the necessary theory is supposed to hold, or the existence of a substitute condition that reflects the underlying concept (see example of Figure 1 (bottom) where GRE score and "opinion of faculty member" both indicate a student's quantitative ability). The evaluation of cases near the ceiling line is not only desirable for reasons of internal validity; it also gives insight in

"peers", "best practices", or "most efficient" cases. Cases near the ceiling line have the highest outcome in comparison with other cases with the same level of the necessary condition. These cases are therefore most efficient in the sense that they have the highest possible outcome for the given level of the necessary condition (outcome efficiency), and the lowest possible level of the necessary condition for the given outcome (condition efficiency).

NCA is a data analysis tool for calculating ceiling lines, effect sizes, and other relevant quantities. If the researcher is only interested in the specific dataset, without a need for generalization to a wider population, this is perfectly acceptable. Also, if the dataset is a census (all members of the population are in the dataset), the calculated quantities are true representations of the population. However if, for reasons of statistical inference, the dataset is a probability sample from a population, the NCA quantities are only point estimates. Currently NCA does not take sampling error into account (e.g., the NCA estimates have no confidence intervals). Future research could focus on adding confidence intervals to point estimates of NCA parameters. Building on developments within the field of production frontiers estimation, such confidence intervals could be based on resampling approaches such as jackknifing and bootstrapping (e.g., Simar & Wilson,1998), or on Bayesian approaches (e.g., Pendharkar & Pai, 2013). Another sampling related limitation is introduced if purposive sampling is applied. Purposive sampling is an efficient way to select samples for calculating NCA-quantities (see the example of selecting successful cases in Appendix 1). The intention is that the sample is representative only for cases that are around the ceiling line (relatively successful cases). Then, relatively unsuccessful cases are not well represented in the sample, and therefore general sample descriptives (e.g., mean values of variables) or average sample trends (regression) do not represent the whole population. It should also be noted that a necessary condition may be trivial. When in a dataset, obtained by purposive or random sampling, a necessary condition is identified, it is possible that this condition is trivial. Trivialness is the extent to which observations, on average, are clustered towards the maximum value of the necessary condition, hence, low values of the condition are rare. For example if the required GRE score for admission would be for example 200, virtually all students would reach that score. Then such score would still be necessary for admission, but it would be trivial because a score below 200 hardly exists. Similarly, references are necessary for an impactful paper in ORM, but this condition is trivial because nearly all papers have references. A check on trivialness requires the identification of the existence of cases without the necessary condition (Dul, et al. 2010).

Recommendations for Applying NCA

NCA can be applied in any branch of organizational research (and elsewhere) in which theoretical necessary (but not sufficient) statements are (or can be) formulated. Table 1 gives examples of such statements in strategy, organizational behavior, and human resource management. Also many examples exist in other organizational research fields such as operations management (Dul et al. 2010; McLachlin,1997; Van der Valk & Wynstra, 2012; Van der Valk et al., 2015) and business history (Westerhuis, Dul, De Jong, & Hak, 2012), in other social sciences including political science (Goertz, 2003), and even beyond that in medical and technical sciences[27]. The examples show that, depending on the focal unit of the theory, necessary conditions can be formulated at any organizational level: individual, team, project, work-unit, business unit, firm, country, etc. What is common in these formulations is that the necessary condition is a non-trivial characteristic, event, resource, or effort, that is relatively unique, scarce, or costly, and that must be designed, controlled, or managed to bring or keep it in place, in order to allow a certain desired outcome to occur. Without the necessary condition there is guaranteed failure of that desired outcome. One may expect to find necessary conditions in datasets with variables that are plausibly causally related, because it seems rare that a desired outcome occurs for free, hence without keeping in place a non-trivial, unique, scarce, or costly condition. Therefore I recommend that researchers should always search for necessary conditions in datasets with presumed causal relations between X and Y, and to acknowledge that necessary causes are different from sufficient causes, and that traditional data analysis techniques (correlation, regression) focus on sufficiency, whereas NCA focuses on necessity. The NCA analysis may complement (e.g., precede) traditional analyses.

Just like traditional data analysis techniques, NCA is a data analysis technique and not a data collection and measurement technique. It assumes that the scores of X and Y are reasonably valid (i.e., measurement instruments measure what they intended to measure) and reliable (precise). Current techniques to check the quality of measurements may be applied. It also assumes that the basic requirements to make it plausible that X causes Y are met. Furthermore, the instances (cases) in the database are considered to represent the group of instances (cases) that is of interest to the researcher, i.e., a specific group of instances (cases) with certain characteristics, a representative sample obtained by probability sampling from a population of interest, or a census of a population of interest.

Building on suggestions provided in Dul et al. (2010), I recommend the following:

- Recognize that traditional analysis (correlation, regression, "average line through the middle"), and necessary conditions analysis (ceiling line) are fundamentally different, and equally valid for their own purposes.

- Recognize that the managerial relevance of a necessary condition analysis (without X there is guaranteed failure of Y, and this cannot be compensated by other determinants) might be stronger and more relevant for practical decision making than that of a traditional analysis (more X on average results in more Y).

- Recognize that it is not uncommon that necessary conditions exist in datasets, although effect sizes may vary.

- Recognize that when traditional data analyses do not show strong traditional relationships, there still may be strong necessary conditions (see example of Figure 4).

- Do not formulate theoretical necessary condition statements ("X is necessary for Y", or "X is necessary but not sufficient for Y"), as traditional hypotheses (e.g., "X affects Y"), as was done in the examples in Table 1.

- Test necessary condition hypotheses with NCA, not with traditional analysis (correlation, regression), as was done in the examples in Table 1.

- Formulate results of a traditional analysis in traditional terms, not in terms of necessary conditions, as was done in the examples in Table 1;

- Formulate results of a necessary condition analysis in necessary condition terms, not in traditional terms (see Dul et al., 2010, for examples).

Below I provide some extended recommendations for researchers wanting to use NCA to analyze their data. These recommendations focus on (1) Performing a data analysis with NCA, and (2) Reporting NCA results. Finally I describe a software tool for conducting NCA.

Recommendations for Performing a Data Analysis with NCA using a Stepwise Approach

Tables 3 and 4 show six steps for performing a necessary condition analysis. Table 3 refers to a situation where both X and Y are continuous variables, or discrete variables with a large number (e.g., >5) of variable levels. Then the analysis is illustrated with a scatterplot as in Figures 3, 4, 5 and 6. Table 4 refers to a situation where both X and Y are dichotomous variables, or discrete variables with a small number (e.g., <5) of variable levels. Now the analysis is illustrated with a contingency table as in Figures 1 and 2. Steps 1 and 2 and Steps 5 and 6 are relatively easy in both the "Scatterplot approach" and the "Contingency table approach". Step 3 ("Draw ceiling line") and Step 4 ("Quantify NCA parameters") are relatively easy in the contingency table approach but more complex in the scatterplot

approach. The NCA software tool has been developed (see below) to assist in performing these steps. All steps in the "Contingency table approach" can be performed by "manual/visual" analysis of the contingency table. This is particularly easy when the number of levels of the discrete variables is relatively low. The analysis for situations with dichotomous variables, or discrete variables with a small number (e.g.,<5) can also be done with the NCA software. Then the alternative data visualization with a Cartesian coordinate system (as in the right hand sides of Figures 1 and 2) and the scatterplot approach with the CE-FDH ceiling line technique must be used. This alternative analysis with NCA software (Table 3) gives the same results as the manual/visual contingency table approach (Table 4).

Table 3

Stepwise Approach for Data Analysis with NCA for Continuous and Discrete Variables with a Large Number of Levels ("Scatterplot Approach").

Step 1 Make scatterplot	Step 2 Identify empty space	Step 3 Draw ceiling line (use NCA software)	Step 4 Quantify NCA parameters (use NCA software)	Step 5 Evaluate effect size and accuracy	Step 6 Formulate the necessary condition
a) Make a scatterplot by plotting the observations in an X-Y Cartesian coordinate system with X-axis "horizontal" and Y-axis "vertical", and values increase "to the right" and "upwards".	a) Visually inspect if the upper left hand corner is empty. Consider to allow some exceptions in the "empty space". b) If there is "empty space" continue with step 3; if not, report that a necessary condition is not observed.	a) Draw a line in the upper left hand corner to separate the "empty space" and the full space using Ceiling Regression (CR-FDH) if a straight line can fit the data points around the ceiling line, or Ceiling Envelopment (CE-FDH) if a piecewise linear function can fit the data points, or use another ceiling technique (see Table 2).	a) Ceiling zone: calculate the size of the "empty space". b) Scope: multiply the highest minus the lowest X values, with the highest minus the lowest Y values. c) Effect size: divide Ceiling zone by Scope. d) Accuracy: the number of observations that are not in the "empty space", divided by the total number of observations, multiplied by 100%. e) Condition inefficiency: (see Figure 3 top right) (XHigh – Xcmax)/ (XHigh – XLow)*100%. f) Outcome inefficiency: (see Figure 3 top right) (Ycmin – YLow)/ (YHigh – YLow)*100%.	a) Evaluate if the effect size (d) is theoretically or practically meaningful in the current context. b) Consider using the general benchmark $0 < d < 0.1$ "small effect", $0.1 \leq d < 0.3$ "medium effect", $0.3 \leq d < 0.5$ "large effect", and $d \geq 0.5$ "very large effect". c) Compare the accuracy with the benchmark of 95%. d) If effect size and accuracy are considered large enough continue with Step 6. If not, report results.	a) Formulate the necessary condition in general terms ("in kind") as "X is necessary for Y" or "X is necessary but not sufficient for Y" [a] b) Additionally formulate the detailed necessary condition ("in degree") by formulating the ceiling line (e.g., $Y_c = aX_c + b$) indicating which minimum level of X_c is necessary for a given level of Y_c.

[a] Because normally there are also observations below the ceiling line, the necessary condition is usually not sufficient.

Table 4

Stepwise Approach for Data Analysis with NCA for Dichotomous and Discrete Variables with a Small Number of Variable Levels ("Contingency Table Approach").

Step 1 Make contingency table	Step 2 Identify empty space	Step 3 Draw ceiling line	Step 4 Quantify NCA parameters	Step 5 Evaluate effect size and accuracy	Step 6 Formulate the necessary condition
a) Make a contingency table with X-axis "horizontal" and Y-axis "vertical", and values increase "to the right" and "upwards". b) Write the number of observations in the cells of this table. c) Delete outer rows and outer columns if these are completely empty.	a) Visually inspect if the upper left hand corner is empty. Consider to allow some exceptions in the "empty space". b) If there is "empty space" continue with step 3; if not, report that a necessary condition is not observed.	a) Draw a line in the upper left hand corner to separate the "empty space" and the full space.	a) Ceiling zone: count the number of "empty cells"[a] in the "empty space". b) Scope: count the total number of cells minus the number of cells of a column, minus the number of cells of a row, plus 1. c) Effect size: divide Ceiling zone by Scope d) Accuracy: the number of observations that are not in the "empty space", divided by the total number of observations, multiplied by 100%. e) Condition inefficiency: divide the number of non-empty cells[a] in the upper row minus one, by the total number of cells in that row minus one, and multiply the result by 100%. f) Outcome inefficiency: divide the number of non-empty cells in the most-left column minus one, by the total number of cells in that column minus one, and multiply the result by 100%.	a) Evaluate if the effect size (d) is theoretically or practically meaningful in the current context. b) Consider using the general benchmark $0 < d < 0.1$ "small effect", $0.1 \leq d < 0.3$ "medium effect", $0.3 \leq d < 0.5$ "large effect", and $d \geq 0.5$ "very large effect". c) Compare the accuracy with the benchmark of 95%. d) If effect size and accuracy are considered large enough continue with Step 6. If not, report the results.	a) Formulate the necessary condition in general terms ("in kind") as "X is necessary for Y".[b] b) Additionally formulate the detailed necessary condition ("in degree"): Select the row with level of the desired outcome for which the necessary condition can be formulated. c) Identify the critical cell in that row: the "non-empty" cell with the lowest level of X. d) Formulate the necessary condition: (at least) the level of X of the critical cell is necessary for the selected level of Y.

[a] Note that an "empty cell" with an exception is an "empty cell" and not a "non-empty cell".

[b] To evaluate if the necessary condition is also "not sufficient" it can be observed if there are non-empty cells with lower levels of Y than the level of Y of the critical cell (the critical cell of a row is the "non-empty" cell with the lowest level of X), and that have the same or a higher level of X than the level of X of the critical cell. If yes, the relation is "necessary but not sufficient". If not, the relation is just "necessary".

Recommendations for Reporting NCA Results

In addition to the common standards for reporting research results, specific elements need to be added for describing the details of a necessary condition analysis. When a necessary but not sufficient theoretical statement is made, it is recommended that researchers should consistently formulate the theory to be tested or built as a necessary (but not sufficient) causal theory (e.g., by formulating propositions/hypotheses in terms of "X is necessary (but not sufficient) for Y", or by formulating a research question "to which extend X is necessary but not sufficient for Y"). Contrary to common practice (see Table 1) such theoretical statements should not be formulated (implicitly) as a sufficiency theory (e.g., by formulating traditional general propositions/hypotheses such as "X affects Y" or "X has a positive effect on Y"). In the Methods section of the report, it can be stated that a necessary condition analysis (NCA) was applied to evaluate (test) or formulate (build) a necessary but not sufficient theoretical statement, and it could be specified how the six steps (Table 3 or 4) were performed. In the "scatterplot approach" the specific ceiling technique (e.g., Ceiling Regression or Ceiling Envelopment) needs to be specified. Also specification of the criteria for evaluation of the practical or theoretical significance of the effect size of the necessary condition is required. If specific criteria for the study's specific context are not feasible, the general benchmark for necessary condition effect size may be selected (i.e., $0< d <0.1$ "small effect", $0.1 \le d <0.3$ "medium effect", $0.3 \le d <0.5$ "large effect", and $d \ge 0.5$ "very large effect"), and for a dichotomous classification an effect size threshold of 0.1 could be used. In the Results section of the report, first the contingency table or the scatter plot (including ceiling line) could be provided. Next a "NCA results table" with basic information about the NCA parameters (Step 4 in Tables 3 and 4) could be formulated for each necessary condition. As a minimum this table should include: ceiling zone, scope, effect size, and accuracy. For discrete and continuous necessary conditions, also condition inefficiency and outcome inefficiency could be included. Table 5 shows the NCA results tables of the three illustrative examples in this paper. In these examples the general benchmark for effect size is used to appraise the results (see the stars). If the effect size is considered large enough, the necessary condition can be formulated *"in kind"*: "X is necessary for Y". Subsequently, the ceiling line or bottleneck table can be used to formulate the necessary condition *"in degree"*: "level of X is necessary for level of Y".

Table 5

NCA Result Tables for the Three Illustrative Examples.

	Dichotomous Example [a] (Figure 1)	Discrete Example [b] (Figure 2)	Continuous Example [c] (Figure 3)
	"In kind" formulation: A GRE score is necessary but not sufficient for admission *"In degree"* formulation: A GRE score of at least 620 is necessary but not sufficient for admission success.	*"In kind"* formulation: BU representation is necessary but not sufficient for Cross-business unit collaboration performance *"In degree"* formulation: 1. A high level of BU representation is necessary but not sufficient for a high level of Cross-business unit collaboration performance. 2. At least a medium level of BU representation is necessary but not sufficient for a medium level of Cross-business unit collaboration performance.	*"In kind"* formulation: Ambition is necessary but not sufficient for Sales ability *"In degree"* formulation: At least a level X_c of Ambition is necessary but not sufficient for a level of Y_c of Sales ability, according to $Y_c=0.045X_c+2.53$
Ceiling zone	1	3	98.2
Scope	1	4	549.2
Ceiling line	Piecewise linear	Piecewise linear	$Y_c=0.045X_c+2.53$
Accuracy	99.7%	100%	96%
Effect size	1[***]	0.75[***]	0.18[*]
Condition inefficiency	NA	0%	33%
Outcome inefficiency	NA	0%	46%

[*] $d \geq 0.1$, [**] $d \geq 0.3$, [***] $d \geq 0.5$
[a] N=342, [b] N=12, [c] N=108

Additionally the bottleneck table can be provided, in particular in the multivariate situation, showing which levels of the condition is a bottleneck for certain desired levels of the outcome. Figure 3 (bottom right) and Figure 4 (lower part) give examples of such tables. For convenience to the readers, the researcher can interpret the bottleneck table and the ceiling line as follows. First the researcher can classify the outcome into several classes, for example three classes (low, medium and high outcome) on the basis of theoretical or practical considerations, measurement scale values (e.g., for a 7 point Likert scale low: <3, medium: 3-5, high>5, the statistical distribution of the data (e.g., low: <25th percentile, medium: 25-75th percentile, high>75th percentile, or any other classification) or the structure of the data. Then for each class it can be indicated how many and which levels of determinants are required to allow that outcome. For example in Figure 4, based on the structure of the data, three classes are distinguished. In the low range outcome values (<50%) no determinant is necessary. In the middle outcome values (50-60%) some but not all determinants are necessary and their levels

need to be up to 20% of the maximum observed levels. In the high range with outcome >70% all four determinants are necessary for a high outcome with required condition levels of 15-73% of the maximum observed level. For the highest outcome (100%) all determinants must have a value of least 50% (and one nearly 100%) of the maximum observed value. When the required condition percentage is below 100%, there is condition inefficiency.

A Software Tool for Necessary Condition Analysis (NCA)

A NCA software tool is made available to identify necessary conditions in datasets. The tool was developed to facilitate the process of drawing ceiling lines, calculating NCA parameters and creating bottleneck tables. The software called `nca.R` is a package that runs with open source programming language R (Culpepper & Aguinis, 2011). The software can be obtained freely from [*needs to be specified*].

The NCA package does three main things:

- It draws scatter plots with ceiling lines. The user can select up to eight different ceiling line techniques presented in Table 5. The defaults are CE-FDH (piecewise linear function) and CR-FDH (straight line).
- It calculates the NCA parameters: ceiling zone, scope, accuracy, effect sizes, condition inefficiency, and outcome inefficiency for each selected ceiling technique. The default scope is the empirical scope calculated from the minimum and maximum values of X and Y in the dataset.
- It calculates the values of variables in the bottleneck table to analyze which X is the bottleneck for a given Y.

The output includes scatterplots with ceiling line(s), a table with the NCA parameters (as well as other relevant information), and the bottleneck table. For running the NCA package, the user can change several parameters to fit the software to the specific situation.

Conclusion

This paper presents the "necessary but not sufficient" interpretation of "X causes Y". Such statements are common in the organizational sciences (and beyond), but the traditional correlation and regression based data analysis frameworks and tools do not fit such interpretations. The Necessary Condition Analysis (NCA) methodology provides a solution. It is new in the organizational sciences and is particularly useful in situations with multi-causality in which several known and unknown determinants (e.g., events, characteristics,

resources, efforts) contribute to a desired outcome (e.g., good performance) but none is sufficient. In such situations, the proposed new approach can identify the "critical" (levels of) determinants that are necessary for achieving the outcome, and that must be put and kept in place in order to make it possible to achieve the desired outcome. Therefore, organizational researchers are encouraged to apply NCA methodology to their research and datasets. Scatterplots that show heteroskedasticity, or relatively low average effects of X on Y are no reasons for not applying NCA: in such situations X might be necessary but not sufficient for Y. With the NCA approach and the freely available NCA software, it is relatively easy to perform such analyses. It may bring new insights to organizational phenomena.

References

Abbott, B. (1974). Some problems in giving an adequate model-theoretic account of cause. In: Charles Fillmore, George Lakoff, & Robin Lakoff, (Eds.), *Berkeley Studies in Syntax and Semantics*, Volume I. Berkeley, CA: Department of Linguistics & Institute of Human Learning, University of California, I-1 - I-14.

Aigner, D.J., & Chu, S.F. (1968). On estimating the industry production function. *American Economic Review*, 58, 826-839.

Autio, E., Dahlander, L., & Frederiksen, L. (2013). Information exposure, opportunity evaluation, and entrepreneurial action: An investigation of an online user community. *Academy of Management Journal*, 56(5), 1348-1371.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. Series: International Series in Operations Research & Management Science, Vol. 157, XVI, 352 p.

Bol, D., & Luppi. F. (2013). Confronting theories based on necessary relations: Making the best of QCA Possibilities. *Political Research Quarterly*, 66(1): 39-44.

Box, G.E.P., Hunter, W.G., & Hunter, J.S. (1978). *Statistics for experimenters. An introduction to design, data analysis, and model building*. John Wiley & Sons, Hoboken, N. J.

Braumoeller, B. & Goertz, G. (2000). The methodology of necessary conditions. *American Journal of Political Science*, 44, 844-58.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6): 429-444.

Chung, K.H. (1969). Can we defend subjective performance appraisal? (A comment). *The Academy of Management Journal*, 12(4), 507-510.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Culpepper, S. A., & Aguinis, H. (2011). R is for revolution: A review of a cutting-edge, free, open source statistical package. *Organizational Research Methods*, 14, 735-740.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. New York, NY: Routledge.

Dettmer, H. W. (1998). Breaking the constraints to world-class performance. ASQ Quality Press.

Diestre, L., & Rajagopalan, N. (2011). An Environmental Perspective on Diversification: The Effects of Chemical Relatedness and Regulatory Sanctions. *Academy of Management Journal*, 54(1), 97-115.

Dion, D. (1998). Evidence and inference in the comparative case study. *Comparative Politics,* 30, 127-145.

Dul, J., & Hak, T. (2008). *Case Study Methodology in Business Research*. Oxford, UK: Butterworth Heinemann.

Dul, J., Hak, T., Goertz, G., & Voss, C. (2010). Necessary condition hypotheses in operations management. *International Journal of Operations & Production Management*, 30, 1170-90.

Dul, J. (2015). Identifying single necessary conditions with NCA and fsQCA. Paper presented at the 4th GIKA Conference, Valencia, July, 13–16, 2015

Evanschitzky, H., Eisend, M., Calantone, R.J., & Jiang, Y. (2012). Success factors of product innovation: An updated meta-analysis. *Journal of Product Innovation Management*, 29 (Supplement S1), 21-37.

Finney, S., & Corbett, M., (2007). ERP implementation: a compilation and analysis of critical success factors. *Process Management Journal*, 13(3), 329-347.

Fiss, P. C. (2007). A set-theoretic approach to organizational configurations. *Academy of Management Review*, 32, 1180-1198.

Gans, J., & Stern, S. (2003). *Assessing Australia's Innovative Capacity in the 21st Century*. Intellectual Property Research Institute of Australia, The University of Melbourne, Melbourne, Australia.

Gilbert, S.F. (2000). Developmental biology. 6th edition. Sinauer Associates, Inc. Sunderland, MA.

Goertz, G. (2003). The substantive importance of necessary condition hypotheses. Pp. 65-94 in *Necessary Conditions: Theory, Methodology, and Applications*, edited by G. Goertz and H. Starr. New York: Rowman & Littlefield.

Goertz, G. (2006a). *Social science concepts: A user's guide*. Princeton, NJ: Princeton University Press.

Goertz, G. (2006b). Assessing the trivialness, relevance, and relative importance of necessary or sufficient conditions in social science. *Studies in Comparative International Development*, 41, 88-109.

Goertz, G., Hak, T., & Dul, J. (2013). Ceilings and floors where are there no observations? *Sociological Methods & Research*, 42(1), 3-40.

Goertz, G. & Mahoney, J. (2012). A tale of two cultures: qualitative and quantitative research in the social sciences. Princeton, USA: Princeton University Press.

Goertz, G. & H. Starr (Eds.). (2003). *Necessary Conditions: Theory, Methodology, and Applications*. New York: Rowman & Littlefield.

Hak, T., Jaspers, F.P.H., & Dul, J. (2013). The analysis of temporally ordered configurations: Challenges and solutions. In P.C. Fiss, B. Cambré, & A. Marx (Eds.), *Configurational Theory and Methods in Organizational Research* (Research in the Sociology of Organization, 38) (pp. 109-127). Bingley, UK: Emerald.

Harvey, F.P. (2003). Practicing coercion: Revisiting successes and failures using Boolean logic and comparative methods. Pp. 147-177 in: *Necessary Conditions: Theory, Methodology, and Applications*, edited by G. Goertz and H. Starr. New York: Rowman & Littlefield.

Hausknecht, J.P., Hiller, N.J., & Vance, R.J. (2008). Work-Unit Absenteeism: Effects of Satisfaction, Commitment, Labor Market Conditions, and Time. *Academy of Management Journal*, 51(6), 1223-1245.

Höfer, T., & Przyrembel, H. (2004). New evidence for the theory of the stork. *Paediatric and perinatal epidemiology*, 18(1), 88-92.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage Publications.

Hogan, R., & Hogan, J. (2007). *Hogan Personality Inventory manual* (3rd ed.). Tulsa, OK: Hogan Assessment Systems.

Hume, D. (1777). *An enquiry concerning human understanding*. London.

James, L.R., Mulaik, S.A., & Brett, J.M. (1982). *Causal Analysis: Assumptions, Models, and Data*, Beverly Hills, CA: Sage.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, 17, 137-152.

Kirca, A.H., Hult, G.T.M., Roth, K., Cavusgil, S.T, Perryy, M.Z., Billur Akdeniz, M, …White R.C. (2011). Firm-Specific Assets, Multinationality, and Financial Performance: A Meta-analytic Review and Theoretical Integration. *Academy of Management Journal*, 54(1), 47-72.

Klayman, J. & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing, *Psychological Review*, 94(2), 211-228.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

LeBreton, (2014). Editorial. *Organizational Research Methods*, 17, 113.

Lewis, D. (1973). Causation. *Journal of Philosophy* 70(17), 556-567.

Madsen A.M., Hodge, S.E., & Ottman, R. (2011). Causal models for investigating complex disease. I. A Primer. *Human Heredity*, 72, pp. 54-62.

Mahoney, J. (2007). The elaboration model and necessary causes. In: Goertz, G., & Levy, J. S. (Eds.). Explaining war and peace: Case studies and necessary condition counterfactuals. New York: Routledge, pp. 281-306.

Mandel, D.R., & Lehman, D.R. (1998). Integration of contingency information in judgements of cause, covariation, and probability. *Journal of Experimental Psychology*, General 127(3), 269-285.

Mandel, D. R., & Vartanian, O. (2009). Weighting of contingency information in causal judgement: Evidence of hypothesis dependence and use of a positive-test strategy. *Quarterly Journal of Experimental Psychology*, 62, 2388-2408.

Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior*, 27, 1031–1056.

McLachlin, R. (1997). Management initiatives and just-in-time manufacturing. *Journal of Operations Management*, 15(4), pp. 271-92.

Marques, J.O.A. (2011). Regularity and counterfactuality in Hume's treatment of causation, *Kriterion: Revista de Filosofia*, 52(124), 355-364.

Martin, J.A., & Eisenhardt, K.M. (2010). Rewiring: Cross-business-unit collaborations in multibusiness organizations. *Academy of Management Journal*, 53(2), 265-301.

Mill, J.S. (1843). A system of logic: ratiocinative and inductive: being a connected view of the principles of evidence and the methods of scientific investigation. London: John W. Parker, West Strand.

Most, B.A., & Starr, H. (2003). Basic logic and research design: Conceptualization, case selection, and the form of relationships. Pp. 25-45 in: *Necessary Conditions: Theory, Methodology, and Applications*, edited by G. Goertz and H. Starr. New York: Rowman & Littlefield.

Pendharkar, P.C., & Pai, D.R. (2013). A Bayesian approach for estimating confidence intervals for DEA efficiency scores under certain inefficiency distribution assumptions. Proceedings for the Northeast Region Decision Sciences Institute, p. 897.

Pizzi, C. (1993). Causality and the transitivity of counterfactuals. *O que nos faz pensar*, 7, 89-103.

Ragin, C.C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.

Ragin, C.C. (2003). Fuzzy-set analysis of necessary conditions. Pp. 179-196. In: *Necessary Conditions: Theory, Methodology, and Applications*, edited by G. Goertz and H. Starr. New York: Rowman & Littlefield.

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*. Chicago: Chicago University Press.

Rihoux, B, Alamos, P, Bol, D. Marx, A., & Rezsohazy, I. (2013). From niche to mainstream method? A comprehensive mapping of QCA applications in journal articles from 1984 to 2011. *Political Research Quarterly*, 66(1): 175-184.

Rowe, A. (1944). *Elimination Diets and the Patient's Allergies*. 2nd Edition. Lea & Febiger, Philadelphia.

Rubin, R.S., Munz, D.C., & Bommer, W.H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, 48(5), 845-858.

Schneider, C.Q., & Rohlfing, I. (2013). Combining QCA and process tracing in set-theoretic multi-method research. Appendix A: A short introduction to the key features of QCA. *Sociological Methods & Research*, March 22. On line appendix: http://smr.sagepub.com/content/early/2013/03/21/0049124113481341/suppl/DC1

Schüz, E. (1940). Bewegungen im Bestand des weissen Storches seit 1934, *Ornithologische Monatsberichte*, 48, 104-120.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Simar, L., & Wilson, P.W. (1998). Sensitivity analysis of efficiency scores: How to bootstrapping nonparametric frontier models. *Management Science*, 44(11), 49-61.

Spector, P.E., & Brannick, M.T. (2010). Common method issues: an introduction to the feature topic in organizational research methods. *Organizational Research Methods*, 13, 403–406.

Stehbens, W.E. (1987), On the "cause" of tuberculosis, *Pathology* 19, 115-119.

Stone-Romero, E. F. (2002). The relative validity and usefulness of various empirical research designs. In S. G. Rogelberg (Ed.), Handbook of research methods in industrial and organizational psychology (pp. 77-98). Malden, MA: Blackwell.

Tulkens, H. (1993). On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts, and urban transit. *Journal of Productivity Analysis*, 4(1), 183-210.

Vaisey, S. (2009) QCA 3.0: The "Ragin Revolution" continues. *Contemporary Sociology* 38(4), 308-312.

Van der Valk, W., Sumo, R.A.F., Dul, J., & Schroeder, R. (2015). How Contracts and Trust are Both Necessary for Innovation in Buyer-Supplier Relationships. Working paper, Tilburg University.

Van der Valk, W., & Wynstra, F. (2012). Buyer–supplier interaction in business-to-business services: A typology test using case research. *Journal of Purchasing and Supply Management*, 18(3), 137-147.

Van Maanen, H. (2010). Worden baby's door ooievaars gebracht? [Are babies brought by storks?]. *Nieuw Archief voor Wiskunde*, 5(11), 134-135.

Vis, B., & Dul, J. (2015). Analyzing relationships of necessity not just "in kind" but also "in degree": Complementing fsQCA with NCA. Working paper.

Walboomers J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X, Kummer, J.A., Shah, K.V., … Muñoz, N. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, 189, 12-19.

Wasserman, E.A., Dorner, D.D., & Kao, S.F. (1990). Contributions of specific cell information to judgements of interevent contingency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16(3), 509-521.

Westerhuis, G., Dul, J., De Jong, A., & Hak, T. (2012). Success in bank mergers. An application of Necessary Condition Analysis (NCA) in business history. Abstract. Proceedings of the 16th annual conference of the European Business History Association, and first joint conference with the Business History Society of Japan, "Business enterprises and the tensions between local and global", Paris, August 30 – September 1, 2012.

Young, R., & Poon S. (2013). Top management support- almost always necessary and

    sometimes sufficient for success: Findings from a fuzzy set analysis. *International*

    Journal of Project Management, 31, 943-957.

Appendix 1: Necessary ingredients of a Caesar salad

Imagine a novice cook planning to prepare a classic Caesar salad[28]. A quick look at recipes reveals that such a salad must have Parmesan cheese. A salad without Parmesan is not a Caesar salad. But a salad with Parmesan does not make it a Caesar salad; Parmesan is necessary but not sufficient for a salad being classified as a Caesar salad. Other ingredients (croutons, romaine lettuce) are necessary as well, but ingredients such as chicken or bacon are not always added, and just contribute to some Caesar salads. A single necessary ingredient, or the set of identified necessary ingredients are not sufficient for the outcome, but they increase the probability of success. However, not using the necessary ingredients will result in guaranteed failure.

How can a novice cook find out which ingredients define a Caesar salad? A traditional approach to answer this question implies an analysis of a large sample of salad recipes in which ingredients are determinants, and a Caesar salad or not is the outcome. A Google search on internet on January 10, 2014 with the keywords "recipes" and "salad" resulted in more than 150 million hits. The 82nd hit was the first Caesar salad. Hence a large N would be required to perform a regression analysis for estimating the contribution of the observed ingredients to the outcome of yes/no Caesar salad. The NCA approach, on the other hand, can do purposive sampling of cases on the basis of the presence of the outcome (Caesar salad). It evaluates which ingredients are present in these cases. Using the keywords "recipe", "Caesar" and "salad" results in more than 3 million hits, and for simplicity I analyzed only the first 10 hits, which all were recipes of Caesar salads. This resulted in the main ingredients presented in Table 6.

It turns out that Romaine lettuce, Croutons and Parmesan are present in these recipes of Caesar salad. Hence these ingredients are candidate necessary conditions: the outcome (Caesar salad) is not possible without them. Anchovy is included in nearly all Caesar salads except for case 6 where it is optional, hence it is virtually necessary. The other ingredients (bacon, chicken) are not necessary: the outcome (Caesar salad) is possible without them. Having the identified necessary ingredients in place will result in a larger probability that the outcome will be a Caesar salad. This can be illustrated as follows. Entering the words "recipe" and "salad" together with one necessary condition (Romaine lettuce, Croutons or Parmesan) in a Google search resulted for the first 10 hits in 0, 1 and 0 recipes of Caesar salad, respectively; hence the average probability that with one necessary ingredient in place a salad is a Caesar salad is $1/30 = 0.03$. When two ingredients are in place the search with the three

Table 6

Main Ingredients of a Caesar Salad (Recipes on Internet).

Case number (hit)[1]	Bacon[2]	Chicken	Anchovy	Romaine lettuce[3]	Croutons	Parmesan cheese[4]
1		x	x	x	x	x
2			x	x	x	x
3			x	x	x	x
4	x		x	x	x	x
5			x	x	x	x
6			x[5]	x	x	x
7			x	x	x	x
8		x	x	x	x	x
9	x	x	x	x	x	x

[1] One hit was deleted because it comprised a large set of cases including case 5
[2] Including speck and prosciutto
[3] Also called cos lettuce
[4] Accepted alternatives: Reggiano or Grana Padano
[5] Optional

possible combinations of two necessary ingredients resulted in an average probability of 0.33 that the salad is a Caesar salad, and with all three ingredients in place the probability that recipes with this set of necessary ingredients results describe a Caesar salad is 0.90 (9 of the 10 first hits were Caesar salads). When including anchovy as a "virtually necessary ingredient" (when allowing some stochasticity) all first 10 hits were Caesar salads (the 13th hit was not a Caesar salad). This example shows that it is relatively easy to find necessary but not sufficient conditions in datasets by focusing on the conditions that are common in cases with high outcome, and that the probability of the high outcome increases when more necessary conditions are in place, but the set of identified necessary conditions is not automatically sufficient for the outcome.

Appendix 2: Necessary conditions for successful business unit collaboration

Martin and Eisenhardt (2010) analyzed success of collaboration between business units in multi-business firms. Table 7 summarizes the data of the twelve cases (collaborations) in this study. From interviews with corporate and business unit informants, the authors obtained discrete (mainly dichotomous and trichotomous) scores for collaboration performance (variables P_1-P_2) and presumed determinants of collaboration success (variables D_1-D_{11}).

Table 7

Data Matrix on Determinants of Collaboration Performance of Twelve Collaborations in Multi-business Firms (Adapted from Martin & Eisenhardt, 2010).

	Performance [a]			Determinants [b]										
				Preparation		Decision					Implementation			
Case	P_1	P_2	P_3	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}
1BH	H	10	H	H	3	H	H	H	H	H	H	H	H	H
2AuH	H	9	H	H	2	H	H	H	M	H	H	H	H	H
3SH	H	9	H	H	2	H	H	H	H	H	H	H	H	H
4VH	H	9	H	H	2	H	H	H	H	M	H	L	H	H
5DH	H	8	H	H	3	H	H	H	H	H	H	H	H	H
6AdH	H	7	M	H	4	L	H	H	H	H	H	H	H	H
7SL	L	6	M	L	0	L	M	L	M	M	M	L	L	L
8AdL	L	4	M	L	0	L	L	L	L	M	H	L	M	L
9DL	L	4	M	L	0	L	L	L	L	M	H	L	L	L
10AuL	L	2	L	L	0	L	L	L	H	L	L	H	H	L
11BL	L	2	L	L	0	L	L	L	M	L	H	L	M	L
12VL	L	2	L	L	0	L	L	L	H	M	H	L	M	L

[a] P_1: Initial original authors' rating of collaboration performance: H= High; L= Low
P_2: Corporate executive rating of collaboration performance: 0=unsuccessful; 5= moderately successful, 10= highly successful
P_3: Trichotomous classification of performance measure P2 by author of current paper: H=High; M=Medium; L= Low

[b] D_1: Origin of collaboration: H=Business unit; L=Corporate
D_2: Number of deliberate learning activities prior to collaboration decision
D_3: Who decided: H=Multi-business team; L=Corporate executive
D_4: General Manager participation: H=High; M=Medium; L= Low
D_5: Choice process: H=Consensus; L=Corporate directive
D_6: General manager ends agreement: H=High; M=Medium; L= Low
D_7: General manager means disagreement: H=High; M=Medium; L= Low
D_8: Business unit representation in collaboration team: H=All; M=some; L=no.
D_9: Full-time team leader: H=yes; L= no
D_{10}: Full-time dedicated team: H=yes; M=some; L=no
D_{11}: Loose coupling: H=yes (3-5 coordinating mechanisms); L= no

The table shows, for example, that determinant "origin of the collaboration" (determinant D_1) is scored at two levels (H=business unit, L=corporate), and that determinant "business unit representation", which is the extent to which all relevant business units are represented in the collaboration team (determinant D_8) is scored at three levels: H=high representation, M= medium representation, and L= low representation. For their purposive sampling strategy, the authors initially scored the outcome variable (performance in terms of collaboration success) dichotomously (H=high-performing, L=low-performing) based on preliminary interviews with a firm's corporate executive (performance measure P_1). Afterwards, collaboration performance was rated by corporate executives using an 11-point scale from 0 = unsuccessful, via 5 = moderately successful to 10 = highly successful (performance measure P_2). For illustration purposes, I trichotomized these performance scores such that scores 0,1,2,3 reflect "Low" collaboration performance, scores 4,5,6,7 "Medium" performance, and scores 8,9,10 "High" performance (performance measure P_3). Table 8 (middle column) shows potential necessary but not sufficient conditions that can be easily derived from Table 7 by visual inspection: looking at the common determinants in successful cases. These hypotheses were not identified by the authors. Instead they derived traditional sufficiency-based hypotheses, namely that the determinant "is more likely to yield" a high outcome (Table 8, right column).

Table 8

Alternative ("Necessary But Not Sufficient"), versus Traditional ("Sufficiency", i.e. "more likely to yield") Propositions Derived from the Data Matrix Presented in Table 7. (After Martin & Eisenhardt, 2010).

Proposition	Alternative propositions based on NCA	Propositions formulated by Martin & Eisenhardt (2010)
1 (D_1)	Business-unit origin (rather than corporate origin) **is necessary but not sufficient** for high-performing cross-business unit collaboration.	"Proposition 1. Business-unit origin **is more likely to yield** a high-performing cross-business-unit collaboration than is corporate origin."
2 (D_2)	Deliberate learning activities conducted prior to a collaboration decision (at least 2) are **necessary but not sufficient** for a high-performing cross-business unit collaboration.	"Proposition 2. Deliberate learning activities conducted prior to a collaboration decision **are more likely to yield** a high-performing cross business-unit collaboration than are other activities."
3 (D_3)	A multibusiness-team decision (rather than a corporate decision) is **necessary** for a high-performing cross-business unit collaboration.	"Proposition 3. A multibusiness-team decision (i.e., one with high GM participation and high consensus) **is more likely to yield** a high-performing cross-business-unit collaboration than is a corporate decision."
4 (D_4)	A high level of General Manager participation is **necessary but not sufficient** for a for a high-performing cross-business unit collaboration.	
5 (D_5)	Consensus decision making (rather a corporate directive) is **necessary but not sufficient** for a high-performing cross-business unit collaboration.	
6 (D_6)	At least a medium level of General Manager ends	

	agreement is **necessary but not sufficient** for a high-performing cross-business unit collaboration.	
7 (D$_7$)	At least a medium level of General Manager means disagreement is **necessary but not sufficient** for a high-performing cross-business unit collaboration.	
8 (D$_8$)	8a. A high level of BU representation is **necessary but not sufficient** for a high-performing cross-business unit collaboration. 8b. At least a medium level of BU representation is **necessary but not sufficient** for a medium level of cross-business unit collaboration performance.	"We assessed the presence of reconfigured teams by whether (1) a specific group was assigned to implement a collaboration, (2) that group included at least one person from every relevant BU who was transferred on a full-time basis to the group, and (3) the group had a full-time leader to whom all members reported." "Proposition 4. A reconfigured team with a few coordination ties **is more likely to yield** a high-performing collaboration than are other types of teams."
9 (D$_{10}$)	A full-time dedicated team **is necessary but not sufficient** for a high-performing cross-business unit collaboration.	
10 (D$_{11}$)	Loose coupling **is necessary but not sufficient** for a high-performing cross-business unit collaboration.	
11	Business-unit origin, deliberate learning activities, multibusiness-team decisions are **temporally ordered necessary but not sufficient** conditions for a high-performing cross-business unit collaboration (details about temporally ordered necessary conditions can be found in Hak, Jaspers, & Dul, 2013).	"Rewiring begins with BU members originating potential cross-BU collaborations. Such a BU origin then triggers deliberate learning activities that provide new experience-based information that helps GMs and other BU executives to decide whether and how to collaborate. If the collaboration looks promising, these deliberate learning activities then set the stage for a multibusiness team decision by: (1) providing shared information about the specifics and value of the collaboration and (2) creating agreement among GMs about the need for the collaboration. This agreement on ends is essential because it motivates very busy GMs to personally take time to move the collaboration forward. But since no single GM has the authority to force collaboration, the GMs are compelled to enter a participative, consensual negotiation in order to proceed. Once the GMs have decided on collaboration, since they are highly committed to it, they are willing to engage in disruptive change, including giving up control and their own employees, to ensure collaboration success. Implementation is accomplished via a reconfigured team with a few coordinating ties. Overall: Proposition 5. Rewiring—a business-unit-centric process—**is more likely to yield** a high-performing collaboration than is a corporate-centric process."

Endnotes

_____

[1] By using *"Or, in other words"* Hume suggests that the first sentence is equivalent to the second one, which may indicate that also Hume confused necessity and sufficiency. Hume's intentions regarding this "two causal definitions" issue, is subject to much philosophical debate (e.g., Marques, 2011).

[2] In Appendix 1 the difference between necessity and sufficiency is further illustrated with an everyday example of the necessary but not sufficient ingredients of a Caesar salad. This example also shows that an NCA analysis for finding necessary ingredients is straightforward.

[3] This expression presumes that all X_i's are necessary and jointly sufficient. In subsection Continuous Necessary Condition the inequality $Y \leq f(X)$ is introduced to represent the necessary but not sufficient condition.

[4] No (inter)disciplinary convention seems to exist regarding the direction of axes in causality contingency matrices. In most studies, the X axis is "horizontal" and the Y axis is "vertical", but exceptions exist (Mandel &Vertanian, 2009). In political science, higher values (from absent to present) are upwards/to the right (e.g., Most & Starr, 2003; Ragin, 2003) or downwards/to the right (e.g., Braumoeller & Goertz, 2000; Harvey, 2003). In psychology of causal reasoning, higher values are usually upwards/to the left (e.g., Klayman & Ha, 1987; Mandel & Lehman, 1998). In organizational sciences, Dul et al. (2010) use higher values upwards/to the right, which I use as the convention in this paper because of its similarity to conventions of Cartesian coordinate systems. I presume that X is the condition that needs to be managed (e.g., event, characteristic, resource, effort) and that Y is an outcome that is desired (e.g., good performance).

[5] This approach corresponds to the Method of Agreement in John Stuart Mill's classic "A System of Logic" (Mill, 1843).

[6] An empty space in the upper left corner corresponds to "X is necessary for Y". An empty space is in the upper right corner corresponds to "-X is necessary for Y". An empty space in the lower left corner corresponds to "X is necessary for -Y". An empty space is in the lower right corner corresponds to "-X is necessary for -Y". This last statement is equivalent to X is sufficient for Y ("floor").

[7] This example introduces the notion of a "virtually necessary condition" (Goertz & Mahoney, 2012). The reason is that in practice exceptions may exist, such that truly deterministic interpretations of necessary conditions do not apply. Dion (1998), Ragin (2000) and Dul and Hak (2008) refer to this phenomenon as "probabilistically necessary", "almost always necessary" or "usually necessary", and "pragmatic determinism", respectively. Thus, when a necessary condition is not in place there is "guaranteed" failure, by which I mean that in practice there is "almost always" or "usually" failure.

[8] The general expression for the number of levels of a discrete determinant constructed from separate discrete items is: $L = L_{max} - L_{min} + 1$, with $L_{max} = \sum_{i=1}^{N} N_{max}(i)$ and $L_{min} = \sum_{i=1}^{N} N_{min}(i)$, where L = number of levels of the constructed determinant, L_{max} = highest level of the constructed determinant, L_{min} = lowest level of the constructed determinant, N = number of items from which the determinant is constructed, i = item number, $N_{max}(i)$ = highest level of the i-th item, $N_{min}(i)$ = lowest level of the i-th item.

[9] If theory or data permits, the piecewise linear function could also be a decreasing function. In this paper, the focus is on (piecewise) linear, non-decreasing ceiling lines.

[10] It is assumed that X and Y have lower and upper limits, hence, cannot be infinite.

[11] If data permits, the line could also be non-linear or a decreasing function. In this paper, the focus is on (piecewise) linear, non-decreasing ceiling lines.

[12] Data obtained from Jeffrey Foster, Vice President of Science, Hogan Assessments, USA, July 9-11, 2014.

[13] The multidimensional ceiling combines the two-dimensional ceiling lines of the separate conditions and is obtained by taking the minimum ceiling value Y_c for a given combination of the four conditions X_1, X_2, X_3, and X_4, thus: the ceiling point Y_c for point X_1, X_2, X_3, X_4 in the multidimensional space is the minimum of [($Y_{1c} = 0.045 X_1 + 2.530$), ($Y_{2c} = 0.022 X_2 + 3.347$), ($Y_{3c} = 0.046 X_3 + 3.161$), ($Y_{4c} = 0.033 X_4 + 3.301$)]. If necessary

conditions are correlated, multivariate NCA will identify each as necessary, although correlated necessary conditions may have the same underlying meaning. In the GRE example, several admission criteria may be correlated (e.g., the GRE score, the TOEFL-score), but each condition is individually necessary and can produce failure.

[14] Effect size names and symbols in the literature are inconsistent. Cumming (2012) discusses names and symbols for "Cohen's *d*" effect size of the difference between means. Cohen (1988) originally used the symbol *d* for *population* effect size, whereas currently most researchers use *d* for *sample* effect size. I follow Cumming's (2012) approach to use *d* for *sample* effect size and reserve δ for *population* effect size, although my *d* for necessary condition effect size is unrelated to "Cohen's *d*" for the effect size of the difference between means.

[15] When the ceiling line is linear, the maximum effect size is 0.5. Only for piecewise linear ceiling lines (e.g., CE-FDH) and for non-linear ceiling lines (e.g., power function $Y=X^{10}$) the maximum effect size can approach 1.

[16] In the multivariate case, the decision rules for rejecting "in kind" necessary condition hypotheses applies to each separate necessary condition. Although it is uncommon in the organizational sciences to formulate "in degree" hypotheses, "in degree" necessary condition hypotheses could be formulated as "[level] X is necessary for [level] Y", where [level] Y is for example a desired level of performance, and [level X is the minimum level of X that is expected to be necessary for that desired level of Y. In the multivariate situation several of such "in degree" hypotheses could be formulated separately. "In degree" necessary conditions can be tested or induced with the bottleneck technique.

[17] If one wishes to use a contingency table for calculating effect size in terms of number of cells, then the ceiling zone is the number of empty cells in the upper left corner (in Figure 2, left: 3), and the scope is the total number of cells (9) minus the number of cells from one column (3) minus the number of cells in one row (3) plus 1. Hence, the effect size calculated from Figure 2, left is 3/ (9-3-3+1) = ¾.

[18] The value $X=X_{cmax}$ cannot be found with a "line through the middle" regression approach. For example, the X-value of the breakpoint in piecewise linear regression, and the X-value of the optimum point in non-linear regression normally do not coincide with $X=X_{cmax}$ because the breakpoint and optimum point depend on the distribution of observations below the ceiling line.

[19] "Trivialness" of a necessary condition is related to necessity inefficiency. Condition trivialness is the extent to which the observations from a representative sample in the necessary condition zone, on average, are clustered towards the maximum value of the necessary condition in that zone (X_{cmax}). Outcome trivialness is the extent to which observations from a representative sample that are in the necessary condition zone, on average, are clustered towards the minimum value of the outcome in that zone (Y_{cmin}).

[20] In the multivariate analysis presented in this paper, combinations of necessary levels of determinants can be considered as a necessary "AND" configuration. In a necessary configuration, each single determinant is necessary, but the number and levels of necessary determinants can change depending on the level of the outcome. Additionally, similar to mediation, a configuration of necessary conditions may be part of a chain of necessary conditions: X_1 is necessary for X_2, and X_2 is necessary for X_3, etc., and the conditions may need to occur in a specific temporal order for the outcome to exist (Hak, Jaspers, & Dul, 2013).

[21] In QCA, original (raw, base) data for X and Y are transferred ("calibrated") into a fuzzy set membership scores between 0 and 1 (2 scores for "crisp" QCA, and many for "fuzzy set" QCA). Therefore, the theoretical scope always equals 1. The empirical scope can be smaller if the minimum and maximum observed scores are not 0 and 1, respectively. The reference line is the diagonal of the *theoretical* scope.

[22] Goertz (2006b) also considers trivialness as an indication of (non-)importance of a necessary condition. Trivialness is the extent to which observations, on average, are clustered towards the maximum value of the necessary condition (see also subsection Limitations of NCA).

[23] The effect size measure of a necessary condition proposed in the present paper can be compared with Goertz' (2006b) "relevance" measure and Ragin's (2006) "coverage" measure for the importance of a necessary condition. When the ceiling line coincides with QCA's reference line (no inefficiencies) and when there are no observations above the ceiling line ($X_i \geq Y_i$ and therefore QCA's "consistency" = 1) the effect size is 0.5, independent of the distribution of the observations below the ceiling line. However, the relevance and coverage

measures depend on this distribution: relevance $= \frac{1}{N} \sum_{i=1}^{N} \left(\frac{Y(i)}{X(i)} \right)$ and coverage $= \frac{\sum_{i=1}^{N} Y(i)}{\sum_{i=1}^{N} X(i)}$, where N is the number of observations. Relevance and coverage reflect the level of sufficiency of the necessary condition, with values between 0 and 1. A value of 1 indicates that all observations are on the ceiling line and that the necessary condition is also sufficient.

[24] The example is inspired by an article of the Dutch science writer, Hans van Maanen, who analyzed the much-cited statistical textbook example of the spurious relation between number of white storks and the size of the population during the period 1929-1938 (Schüz, 1940; Van Maanen, 2010). He showed that the data presented in this common example are incomplete (which resulted in a more stylized example with fewer scattered data points), and that the selected outcome measure (population) is not compatible with the hypothesis that storks bring babies. In my example, I used the complete dataset received from Van Maanen (January 19, 2014) with the number of newborns as the outcome variable.

[25] A more plausible explanation for the increase of the number of births is provided by Van Maanen (2010). He suggests that this increase is due to the birth encouragement policy of the German regime prior to World War II, and that the increase of stork nesting pairs is a natural fluctuation.

[26] In (genetic) biology the "necessity" experiment is called a "loss-of-function" or "knock-out" experiment, and a "sufficiency" experimental approach is called a "gain-of-function" or "knock-in" experiment (e.g., Gilbert, 2000, p. 43). In medicine, an "elimination diet" (Rowe, 1944) is an example of a "necessity experiment" to discover foods or ingredients that cause food allergy.

[27] For example, in medicine it is accepted that pathogenic microorganisms, such as bacteria and viruses, can be necessary but not sufficient causes of infectious diseases. For instance, the tubercle bacillus is a necessary but not sufficient cause of tuberculosis (Stehbens, 1987), the Human Immunodeficiency Virus (HIV) is a necessary but not sufficient cause of AIDS (Madsen, Hodge, & Ottman, 2011), and the Human Papilloma Virus (HPV) is a necessary but not sufficient cause of cervical cancer (Walboomers et al., 1999).

[28] Although the relationship between ingredients and Caesar salad is ontological (defining the concept, Goertz, 2006a) and not causal, this example easily illustrates the fundamental difference between necessity and sufficiency.