

Roel Rutten*

Comparing causal logics: A configurational analysis of proximities using simulated data

<https://doi.org/10.1515/zfw-2019-0023>

Received: 24. October 2019; accepted: 16 March 2020

Abstract: Unnoticed by economic geography for fifteen years, Boschma's (2005) proximity paper conflates two different causal logics: regularity and substantive interpretation. The former is dominant in variable-based methods, the latter in case-based methods. Using the proximities approach as an example, this paper explains the differences between both logics. A QCA (Qualitative Comparative Analysis) study on simulated data demonstrates how case-based methods use substantive interpretation for causal inference. QCA is an important innovation in case-based methods that, thus far, economic geography has largely missed. QCA challenges the search for causal effects of individual causes and presents configurational causality as a compelling alternative.

Keywords: Causality, critical realism, innovation, proximities, QCA (Qualitative Comparative Analysis)

Introduction

Explaining whether geographical proximity matters for innovation is a key concern for economic geography. Different theories and methods have been used to investigate this question. They often lead to different findings because they follow different causal logics. This paper discusses how theorizing and empirical research differs between variable-based and case-based methods, using the example of Boschma's (2005) paper on proximities and innovation. The paper suggests QCA (Qualitative Comparative Analysis), which is little known in economic geography, as a configurational alternative to the causal-effects approach of variable-based methods. The differences between variable-based and case-based methods can be demonstrated by how they interpret the XY-plot of

Figure 1. This figure shows simulated data on inter-firm dyads; pairs of firms that are collaborating on innovation across degrees of distance. Boschma (2005) investigates how different proximities explain innovation, and whether geographical proximity matters. Figure 1 shows what, following Boschma (2005), the relationship between cognitive proximity and innovation would look like. For a variable-based researcher, i. e. a researcher trained in looking for correlations between variables, the XY-plot suggests a curvilinear effect of cognitive proximity on innovation. The researcher would theorize that higher levels of cognitive proximity between dyad partners cause higher levels of innovation because having overlapping knowledge bases allows firms to learn from one another. But only up to a point; viz. when their knowledge bases overlap too much, firms no longer learn from each other causing innovation levels to deteriorate (Boschma 2005: 63–64). This is probably the way that, intuitively, most researchers would look at the XY-plot. Regardless of whether they use quantitative or qualitative methods, most social scientists look for variation across a population. This variation, or variable-based, logic is often suggested as the overarching logic of causal inference in the social sciences, applying equally to qualitative and quantitative methods (King et al. 1994). However, a researcher following a case-based logic would look at Figure 1 very differently (Goertz 2017: 59–60, Ragin 2008: 23–25).

A case-based researcher would see a number of cases (i. e. dyads) that are both cognitively proximate and innovative (Cell I). This would suggest a causal mechanism connecting cognitive proximity to innovation. However, there are also plenty of cases in Cell II, where the cause (cognitive proximity) is present, but not the outcome (innovation). These cases defeat the causal mechanism suggested by Cell I, indicating that cognitive proximity is not sufficient for innovation; i. e. the presence of cognitive proximity does not explain innovation. There are also plenty of cases in Cell III. They suggest that dyads can be innovative without being cognitively proximate; that is, other reasons (causal mechanisms) than the one suggested by Cell I also explain innovation. Consequently, cognitive proximity is not necessary for innovation. Finally, in case-based research, Cell IV is causally irrelevant. In variable-based research, for a linear effect, these cases contribute equally

*Corresponding Author: Dr. Roel Rutten, Tilburg School of Social and Behavioural Sciences, PO Box 90153, 5000 LE Tilburg, The Netherlands; ERAC (European Regional Affairs Consultants), PO Box 90102, 5200 MA 's-Hertogenbosch, The Netherlands; Newcastle Business School, Northumbria University, City Campus East, Newcastle upon Tyne, NE1 8ST, United Kingdom, e-mail: r.p.j.h.rutten@uvt.nl, telephone: +31 13 466 2164

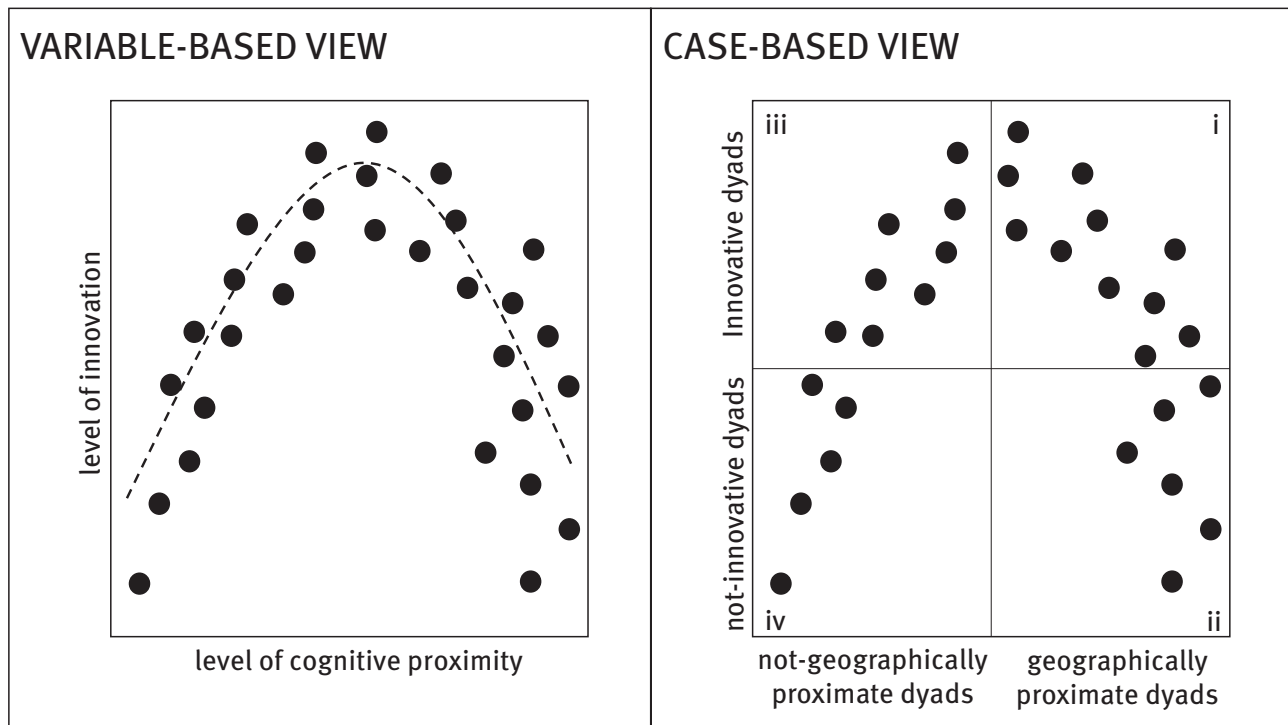


Figure 1

to the correlation coefficient as the cases in Cell I, but since neither the cause nor the outcome is present, these cases tell us nothing about how or why cognitive proximity contributes to innovation¹. So, a case-based researcher would conclude that, whatever causes innovation, it is not cognitive proximity (cf. Goertz and Mahoney 2012a: 19–23).

The same data, two juxtaposed conclusions. Who is right? That depends on whether one follows a variable-based or a case-based causal logic. The former clearly being the dominant approach of mainstream social science (Ragin 2014: 53). The aim of this paper is to introduce the case-based logic of causal inference and to explain how it differs from the variable-based approach. The paper uses Boschma's (2005) proximity approach to illustrate these differences for an economic geography audience. The proximity approach is particularly suited for this purpose because the Boschma (2005)-paper (unintentionally) conflates both logics. On the one hand, it aims to isolate the *effect* of each proximity on innovation (interactive learning, in the paper's language). On the other hand, it wants to investigate which *combinations* of proximities explain innovation. The fact that it is one of the most cited papers

in economic geography is important for two reasons. First, it requires little introduction because economic geographers will be familiar with it. Second, the fact that few (if any) economic geographers have drawn attention to the conflating of two causal languages shows the relevance of the present paper's effort. It suggests that (mainstream) economic geography may learn something new.

Following a case-based causal logic, QCA is one of the most innovative developments in case-based methodology (Greckhamer et al. 2018, Mahoney 2001). QCA continues to gain in popularity across the social sciences but entered economic geography only very recently (Legendijk et al. 2020, Li and Bathelt 2020, Rutten 2019). QCA is a comparative case-study method that calculates cross-case regularities (co-occurrences of causes and outcomes) using Boolean algebra and then interprets them into causal mechanisms using contextual and in-depth case-knowledge. Contrary to, for example, regression analysis, in QCA, the robustness of empirically observed regularities does not evidence causality. The plausibility of the putative cause substantively explaining the outcome is what matters for causal inference in QCA, and case-based methods in general (Ragin 2008, 2014; Schneider and Wagemann 2012). QCA was initially developed for 'mid-size' Ns (10–50 cases); too many for conventional case-based (qualitative) methods and too few for conventional variable-based (quantitative) methods (Schneider

¹ In case-based research, causality is asymmetrical. Things that explain the presence of the outcome are not necessarily the same as the things explaining the absence of the outcome. A separate analysis must be performed to explain the absence of the outcome.

and Wagemann 2012: 10). Nowadays, QCA is increasingly performed on (much) larger datasets². While technically possible, a large-N distances researchers from their cases and thus defeats one of QCA's key strengths. Distance to cases forces researchers to rely on the robustness of empirical findings rather than on substantive interpretation in order to infer causality. For methodological and philosophical reasons that are beyond the scope of this paper, this creates important problems (Emmenegger et al. 2014; Rutten, forthcoming). After discussing the differences between variable-based and case-based causal inference, the paper explains how QCA works using simulated data on proximity and innovation in dyads.

Causality in variable-based and case-based methods

Case-based researchers have profoundly rejected the notion of a common overarching method of causal inference in the social sciences after King et al. (1994) published their book. While they do not agree on the 'correct' method, there is broad agreement that case-based causal methods follow a very different causal logic than do variable-based methods (Beach and Pedersen 2013, Byrne and Ragin 2009, Goertz 2017, Goertz and Mahoney 2012a). (See also Abbott 1988 and 1998 for a comprehensive critique of the variable-based, regularity approach to causality.) Set-analysis is the principle method of causal inference in case-based methods, among others, informing process tracing, for within-case analysis (Beach and Pedersen 2013) and, perhaps most notably, QCA, for cross-case analysis (Ragin 2008, 2014). The variable-based, case-based divide goes beyond the qualitative-quantitative divide. The fundamental difference being that variable-based researchers decompose cases into variables and look for regularities (i. e. correlation/co-variation) between variables. This can be done quantitatively, using regression analysis or other statistical methods, but also qualitatively, for example with pattern matching. Instead, case-based researchers argue that cases are analytically relevant and should be investigated holistically (Easton 2010: 120, Ragin 2014: 13–14, Schneider and Wagemann 2012: 24). Investigating cases relies heavily on qualitative methods but case-based researchers make use of quanti-

tative techniques as well (Bayesian statistics in process tracing, Boolean algebra in QCA).

As Figure 1 shows, plotting the 'level of cognitive proximity' against the 'level of innovation' of many dyads may produce a curvilinear shape. This suggests an effect of cognitive proximity on innovation. That is, the independent variable cognitive proximity 'does something' to the dependent variable innovation. However, that is not what happens in any one dyad at any one point in time. On the level of a case, proximity and innovation between firms does not vary; they are fixed characteristics. So the question how proximity causes innovation in a particular dyad, is not answered by showing a pattern on the level of a population of dyads. That question can only be answered by explaining how and why the presence of proximity makes it possible for dyads to produce innovations. For example, Cell I clearly suggests that, in some cases, cognitive proximity explains innovation. The underlying mechanism may be that partially overlapping knowledge bases allow firms to exchange tacit knowledge. However, Cell II suggests that this is not a general cross-case pattern. Clearly, cognitive proximity alone is not enough for cases in general to achieve innovation. In case-based language, the 'causal power' of cognitive proximity is negated by some other factor(s). This may be the absence of, for example, geographical or social proximity. Without geographical proximity, face-to-face exchange of tacit knowledge may be difficult. Without social proximity, firms may not trust each other to share their tacit knowledge. That is, while cognitive proximity itself does not in general make it possible for firms to be innovative, cognitive proximity in combination with geographical or social proximity does. Case-based causal logic thus suggests that the configuration of cognitive AND geographical proximity, or the configuration of cognitive AND social proximity does, in general, make it possible for firms to be innovative (Bhaskar 2008: 49–50, 231–232; Byrne 2009: 102–103; Collier 1994: 43, 126–129; Kent 2009: 194–195; Ragin 2008: 23–25, 2014: 19–26). If we would make an XY-plot where the cause is the configuration cognitive AND geographical proximity, Cell II would be (nearly) empty of cases.

This example demonstrates that causality in case-based analysis is complex. Complex causality is configurational, equifinal and asymmetrical (Goertz and Mahoney 2012a: 56–58, Ragin 2014: 19–20, Schneider and Wagemann 2012: 56–58). That is, configurations rather than individual causes explain outcomes. Instead, variable-based methods identify effects of individual causes which are then added (e. g. in a regression model). Complex causality is equifinal because multiple configurations may explain the outcome, i. e. there are multiple causal mechanisms.

² One can do 150 interviews and still have in-depth case knowledge and one can take 15 cases from a database and have no in-depth case knowledge. The in-depth case knowledge rather than the number of cases matters.

In variable-based methods, a regression equation suggests one model that explains all variance across the population (unifinal causality). Finally, complex causality is asymmetrical because, when the presence of a cause explains the presence of the outcome, it does not mean that the absence of the cause explains the absence of the outcome. For example, the presence of geographical proximity may explain the presence of innovation in some dyads, but in other dyads geographical proximity may be absent while the dyads are still innovative. Instead, variable-based causality is symmetrical because the same equation explains both high and low levels of the outcome. Thus, in case-based research, whether geographical proximity explains innovation depends on the presence or absence of other causes. This contrasts sharply with the variable-based aim of identifying a general effect of geographical proximity on innovation. (Variable-based researcher will now say: interaction effects! This is addressed below.)

Crucially, variable-based methods predict outcomes, while case-based methods detect causes. That is, they ‘work’ in opposite directions, which has major implications for theorizing and empirical strategies. Looking at Figure 1, a variable-based researcher would argue correctly that, up to a certain point, the higher the level of cognitive proximity, the higher the level of innovation one may expect to find in a firm. That is, variable-based researchers theorize in terms of variables and try to identify empirically robust relations (a regularity) between them to evidence the theorized mechanism. Case-based researchers object that population-level effects say nothing about individual cases: population-level effects are not case-level probabilities (Kent 2009: 188–190, Lawson 2005: 373–376). Case-based researchers further object that variables are not real; they are statistical constructs that do not actually exist (Byrne 2009: 4, Kent 2009: 188, Ragin 1992: 5). Variable-based researchers may not suggest population-level effects as case-level probabilities but the fact remains that they theorize and research variables, not cases. Case-based researchers start by observing that some firms are innovative and then detect which configurations of causes might explain that. Their focus is not on empirical robustness but on substantive interpretation. A cross-case pattern (regularity) does not have to be empirically robust. There may be good reasons why a dyad is not innovative even though it is characterized by the presence of geographical proximity AND cognitive proximity. A recent innovation project may have failed, for example. That is, substantive interpretation (based on in-depth case-knowledge) may lead a case-based researcher to conclude that the configuration of geographical proximity AND cognitive proximity may have the causal power to ‘produce’ inno-

vations but that this does not determine that it will. The causal power needs to be exercised (firms have to actually engage in tacit knowledge exchange) and even then the effort may fail (as in a failed innovation project). Put differently, it is the agents (firms) and their causal powers (reflected in configurations of causes) that are the principal object of case-based analysis (e.g. QCA), and not the propositions about them (i.e. the hypothesized mechanisms connecting independent and dependent variables) (Gorski 2013: 665. See also Bhaskar 2008: 113–118; Ragin 2008: 112, 2014, 16–17). Another way of saying the same thing is that variable-based methods prioritize epistemology while case-based methods emphasize ontology. For example, for variable-based researchers, an empirically robust correlation between cognitive proximity and innovation would evidence a hypothesized causal mechanism connecting them. That is why variable-based researchers use sophisticated statistical techniques to estimate the correct effect of an independent variable. For case-based researchers, empirically observed patterns are purely descriptive. Many would argue that one cannot directly observe a causal mechanism at all; we can only see its observable implications (Bhaskar 2008: 215, Collier 1994: 26–28, Lawson 2005: 380–383, Mahoney 2001: 580). Consequently, through substantive interpretation, researchers must get from empirical observations (epistemology) to the underlying causal mechanisms (causal powers, i.e. ontology). So a causal mechanism, in case-based research, is an analytical (not an empirical) statement (Little 2015: 468, 473–474).

Thinking about causal mechanisms in terms of ontology and causal powers, has a very strong basis in critical realism (Bhaskar 2008, 2014; Collier 1994; Gorski 2013, 2018; Lawson 2005). QCA in particular may be seen as a critical realist method because it clearly distinguishes cross-case empirical patterns (epistemology) from underlying causal mechanisms (ontology) (Gerrits and Verweij 2013; Rutten, forthcoming). Furthermore, establishing when a case counts as, for example, geographically proximate and when not, requires answering two questions: What does it mean when we say that firms are geographically proximate (ontology), and how do we know geographically proximate firms when we see them (epistemology). For example, one way to express geographical proximity is the time it takes to travel to the location of one’s partner. Variable-based researchers would make a scale that ranges from zero hours (co-location) upwards. Case-based researchers, instead, would pick a value below which partners are considered geographically proximate and above which they are not-geographically proximate. Which value to select depends on how geographical prox-

imity is ontologically defined; which, in turn, depends on context. Researching geographical proximity between knowledge workers at a campus, one might argue that less than five minutes walking is geographically proximate. Studying inter-firm collaboration on innovation in dyads, one might set that threshold at a two-hour drive. Ontologically, geographical proximity may be defined as being able to visit one's partner's location, have at least a two-hour meeting, go back and discuss the outcomes of the meeting at one's own location, on the same day. Epistemologically this may be interpreted as no more than a two-hour drive one-way. Thus, while a scale in variable-based research suggests that all differences are equally relevant, case-based researchers distinguish between differences in kind and differences in degree. In this example, a firm relocating from 10 to 8 hours driving from its partner, still counts as not-geographically proximate. A firm relocating to reduce travel time from three hours to one, becomes a geographically proximate firm. The same two hours are a difference in degree in first case but a difference in kind in the second. (See Goertz and Mahoney (2012b) on the ontology and epistemology of measurement.)

Variable-based researchers have an intuitive dislike against 'throwing away' variation and thinking in terms of the presence (1) or absence (0) of causes and outcomes. It makes no sense when one is looking for co-variation between variables. But when the focus is on cases, not variables, and when causality is thought of as the causal power of cases (which follows from them having or not-having certain characteristics), it becomes a different story. Seen this way, much in the world is dichotomous. Either firms trust each other to exchange tacit knowledge, or they do not. Firms are either located in a high-technology region, or they are not. And this has consequences for the causal power of these firms to engage in knowledge exchange. On the level of a population, a linear effect suggests that a low level of cognitive proximity causes a low level of innovation. On the level of cases, exchange of tacit knowledge will not happen until a threshold of cognitive proximity between firms is reached. Moreover, thinking in terms of set-memberships allows refining empirical analysis because sets are asymmetrical. If the threshold for being geographically proximate is a two-hour drive, than firms more than two hours away are 'not-geographically proximate'; which is something qualitatively different than 'geographically distant'. Ontologically, one may speak of a 'geographically distant' firm if it is, e. g., more than a five-hour drive away. This allows a researcher to investigate, for example, whether the absence of 'geographical proximity' or the presence of 'geographical distance' explains the absence of tacit knowledge exchange between firms.

Variable-based interaction versus case-based configurations

Boschma (2005) developed his paper in perfect alignment with the dominant variable-based approach in economic geography. This probably explains a good deal of the paper's popularity; it connects to how many economic geographers think about causality. Boschma (2005) identifies five proximities (geographical, cognitive, social, organizational and institutional proximity) and theorizes how each of them affect innovation. At the same time, Boschma (2005) understands that outcomes are not usually explained by the effects of individual independent variables alone. However, the wording that Boschma (2005) chooses to make the argument is particularly unfortunate because it conflates different causal logics. He begins by arguing that geographical proximity is neither sufficient nor necessary for interactive learning (p. 62). It is not necessary because other proximities may allow interactive learning and it is not sufficient because interactive learning across distance requires other proximities to be effective. The problem with this statement is that it decontextualizes causality. When variable-based researchers identify an empirically robust effect of geographical proximity on innovation, the (implicit) assumption is that this effect holds across contexts. Boschma (2005) makes this very clear when he argues that "it is essential for analytical reasons to ... research and assess the effects of each dimension [of proximity] on interactive learning and innovation. In other words, there is a strong need to *isolate analytically* the effect of geographical proximity from the other forms of proximity to determine whether geographical proximity really matters for the process of innovation" (p.62, emphasis added). Following mainstream, variable-based thinking, Boschma (2005) thus clearly suggests that each proximity (independent variable) has its own effect on innovation and that this effect holds across contexts, i. e. in any innovation process. The latter allows establishing whether geographical proximity "really matters" (in general). However, for case-based researchers, causality is contextual and sufficiency and necessity are not general claims that hold across contexts. In some contexts, geographical proximity may be necessary for innovation. For example, experimenting with a particular prototype requires engineers to be physically in the same place as the prototype. That is, the place-specific nature of some kinds of knowledge (creation) is an important scope condition for the necessity of geographical proximity. Thus, a case-based researcher might argue that geographical proximity is necessary for innovation if knowledge (crea-

tion) is place specific (cf. Rutten 2017: 171–173). In a similar vein, geographical proximity may be sufficient for innovation. For example, geographical proximity is sufficient for knowledge exchange between economic geographers at a conference. No other factors are required. Of course, some level of cognitive proximity exists between them but this may be defined as the absence of ‘cognitive distance’ rather than the presence of ‘cognitive proximity’; because of the diversity of theories and methods in the field. Consequently, a case-based researcher might claim that geographical proximity is sufficient for innovation between members of the same professional community.

At first glance, Boschma (2005) seems to capture this by suggesting to identify “a number of mechanisms [proximities] that ... by their own (sic), or *in combination* cause innovation” (p.62). He further argues that “there is little understanding of the possible *combinations* of the various forms of proximity” and their effect on innovation (p. 62, emphases added). The problem is that ‘combination’ in linear algebra is an interaction but in Boolean algebra it is a configuration. And they are very different things; in terms of both their empirical (epistemological) and conceptual (ontological) implications (Kent 2009: 190–191). This is easily demonstrated by looking at Figure 2. The variable-based approach uses statistics (linear algebra) to estimate the effects of independent variables on dependent variables. The case-based approach uses Boolean algebra to detect which configurations of putative causes explain the outcome of interest. Looking at the left pane of Figure 2, combination clearly means interaction. In this example, a variable-based researcher would argue that:

$$\text{INNOV} = \beta\text{GeoProx} + \beta\text{CogProx} + \beta\text{SocProx} + \beta(\text{GeoProx} * \text{CogProx}) + \beta(\text{GeoProx} * \text{SocProx}) \quad (1)$$

INNOV (innovation),
GeoProx (geographical proximity), CogProx (cognitive proximity), SocProx (social proximity).

Note that the equation suggests simple causality: effects are added, the added effects suggest one causal explanation of innovation (unifinal), and the same equation explains both high and low levels of innovation (symmetry) (Ragin 2014: 82–84). Note also that the equation is fairly complicated so simple does not suggest simplistic. More importantly, note that the effects (the β s) of the interaction terms are independent of the effects (β s) of the individual independent variables. Put differently, linear algebraic interactions are agnostic about whether or not the individual terms in the interaction also have an effect on

the outcome. This is fundamentally different in Boolean algebra.

Look at the right pane of Figure 2 to see the basics of set-theory and Boolean algebra. The outcome of interest is innovation; we want to explain why cases (dyads) are innovative, so we look at the set of innovative dyads (represented by the large circle). One of the putative causes for innovation is geographical proximity, so we also look at the set of geographically proximate dyads (represented by the middle one of the three small circles). All dyads (cases) that qualify as innovative are within the circle for innovative dyads. All dyads that qualify as geographically proximate (viz. less than two hours driving between the firms in the dyad) are in the circle for geographically proximate dyads. Note that a portion of the circle for geographically proximate dyads is outside of the circle for innovative dyads (Area E). This means that some dyads (cases) are geographically proximate and also innovative (Areas A, B, C and D); however, other cases are geographically proximate but not innovative (Area E). These cases are in Cell II of Figure 1 (i. e. Cell II in Figure 1 and Area E in Figure 2 both represent inconsistent cases). Set-theoretically this means that the set of geographically proximate dyads is an inconsistent subset of the set of innovative dyads. Therefore, within the context of this study (example), there is no cross-case pattern connecting geographical proximity and innovation and, consequently, geographical proximity is not sufficient for innovation. Now look at Areas A and C. These are the cases where firms are both geographically proximate and cognitively proximate. In set-theory speak, this area is the intersection of the sets of geographically proximate dyads and cognitively proximate dyads. This is visualized as the circles (sets) for geographical and cognitive proximity partially overlapping (Areas A and C). In Boolean algebra, an intersection is expressed as a logical AND (also written as an *). A logical AND means that both terms in the configuration are present: geographical proximity AND cognitive proximity. Note that all of Areas A and C is within the circle for innovation. That is, cases characterized by geographical proximity AND cognitive proximity are a consistent subset of innovative cases. Consequently, this configuration is sufficient for innovation.

The same goes for the configuration geographical proximity AND social proximity (Areas B and C). So, in this example, there are two explanations for innovation. Geographical proximity AND cognitive proximity might suggest that firms are able to exchange tacit knowledge in face-to-face interaction. Geographical proximity AND social proximity might suggest that co-location allowed the development of strong ties through which tacit knowledge can be exchanged. A Boolean configuration (logical

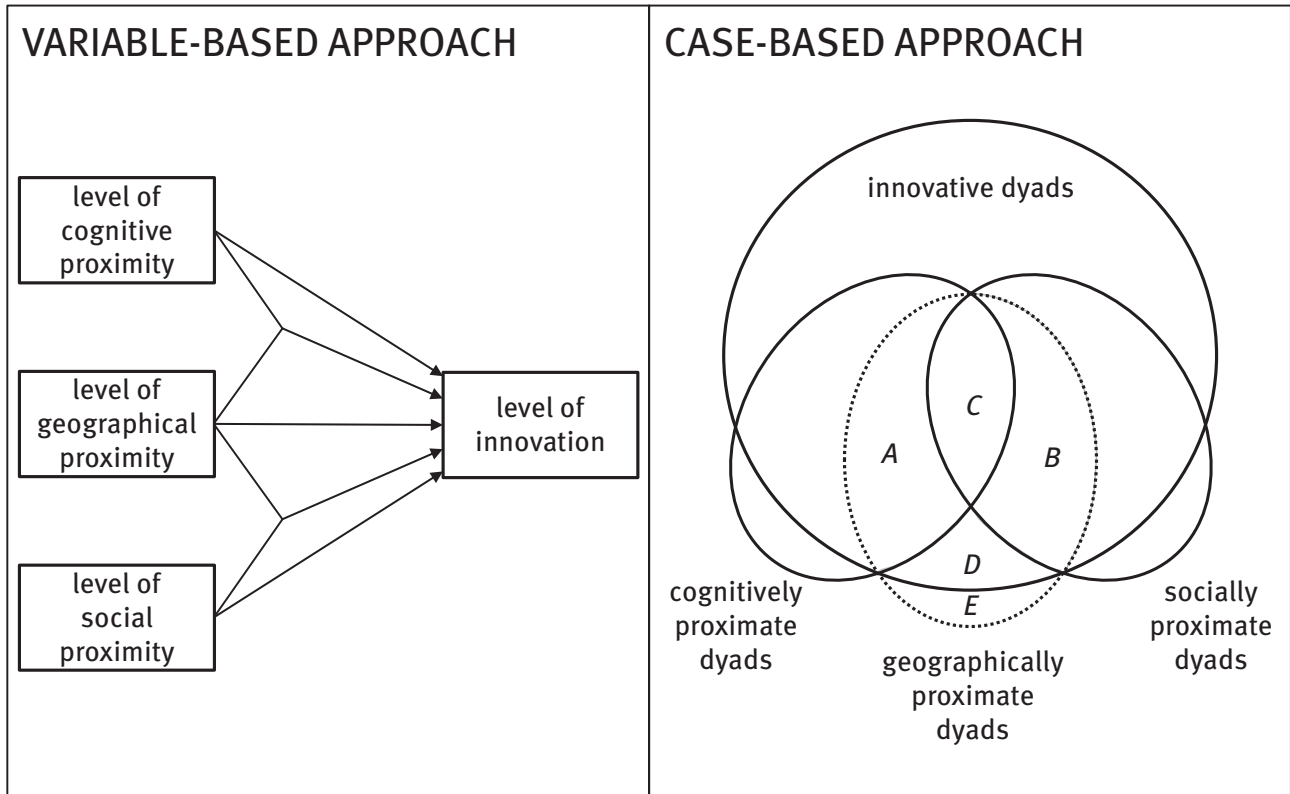


Figure 2

AND) thus rules out a separate ‘effect’ of the individual terms on the outcome. The precise interpretation of the causal mechanisms suggested by the configurations must follow from researchers’ in-depth case knowledge and knowledge of context. Geographical proximity AND cognitive proximity (Areas A and C) and geographical proximity AND social proximity (Areas B and C) are two different causal explanations. Area C suggests that some cases are characterized by geographical proximity AND cognitive proximity AND social proximity. This is not a separate causal mechanism; these cases are ‘overdetermined’. Both causal mechanisms are ‘at work’ in these cases. Areas A and C and Areas B and C together are the union of the two configurations. In Boolean algebra, unions are expressed as a logical OR (also written as a +). This means that, in this example, the ‘full model’ can be written as:

$$\text{INNOV} = (\text{GeoProx} * \text{CogProx}) + (\text{GeoProx} * \text{SocProx}) \quad (2)$$

Which may also be written as:

$$\text{INNOV} = \text{GeoProx} * (\text{CogProx} + \text{SocProx}) \quad (3)$$

Note that the equation expresses complex causality: it is configurational, it shows two causal mechanisms (equifinal), and the equation only explains the presence of innovation, not its absence (asymmetrical). Note also that the equation is fairly simple. The difference between simple and complex causality makes Boschma’s (2005) reference to ‘combinations’ problematic. It conflates two very different ways of theorizing and researching causality. Linear algebraic interactions connect to fundamentally different assumptions on the nature of causality and matching empirical strategies than Boolean algebraic configurations. For a full discussion of set-theory and Boolean algebra, that also includes necessity, see Schneider and Wagemann (2012: 41–90), and Ragin (2014: 85–102).

A QCA example: Proximities and innovation

The most comprehensive QCA handbook available to date is Schneider and Wagemann (2012), best read in combination with the work of Ragin (2008, 2014). The below example focuses on QCA's most recognizable feature: the truth-table analysis. Note that the truth-table analysis is only one step in a QCA study and that the steps before the truth-table analysis (selecting cases and calibrating them into sets) and after (substantive interpretation of the empirical findings) are much more consequential for the quality of a study. The various steps in a QCA study are designed to make optimal use of researchers' case-based and contextual knowledge. Table 1 presents a brief discussion of the key steps (cf. Greckhamer et al. 2018). Because this example focuses on how the truth-table analysis 'works', it uses simulated data. This bypasses a lengthy discussion on case selection and calibration. (See Schneider and Wagemann (2012) and Ragin (2008, 2014) on these issues.) Following QCA conventions, in equations, this section expresses logical ANDs as * and logical ORs as +. Furthermore, the negation of a set, e.g. 'not-geographically proximate', is expressed as ~geographically proximate.

The truth table (Table 2) shows all logically possible combinations of the conditions (causes) under investigation. Each possible combination (configuration) is a row in the truth table. Each condition can be present (1) or absent (0) in a row, meaning that a truth table has 2^x rows, where X is the number of conditions. The number of rows thus increases exponentially with the number of conditions so researchers must limit the number of conditions in their analysis (Schneider and Wagemann 2012: 92, 152–153). This example uses four conditions (cognitive, social, organizational and geographical proximity), which results in a truth table of 16 rows. Also including institutional proximity would have doubled that number to 32 rows, which is not practical for this example. The truth table further shows whether a row is sufficient for the outcome, i. e. whether it explains innovation. If so, the row gets a 1 in the outcome column; if not, a 0. Researchers calibrate (assign set-membership to) their cases in conditions and the outcome based on in-depth case knowledge.

Table 1: Key QCA steps

Step 1: Case selection. Carefully constructing a population of comparable cases. A variable-based researcher would compile a database with as many dyads as possible. A case-based researcher is conscious of the fact that, for example, innovation dynamics in small-firm dyads may be different from large-firm dyads and, consequently, would only select small-firm dyads. In practice, variable-based researchers often work with given populations, and use control variables. Case-based researchers will always construct a population of analytically comparable (homogenous) cases (Goertz and Mahoney 2012 a: 41–48, Ragin 1992: 1–3).

Step 2: Selecting conditions (putative causes) and outcomes. QCA can only handle a limited number of conditions. More than five makes results difficult to interpret. Consequently, the selection of conditions and outcomes is strongly guided by theoretical and contextual knowledge. Moreover, QCA is an ontological method so conditions and outcomes must be defined and measured in a way that makes them causally interpretable. Investigating the relationship between openness and innovation of regions, Gambardella et al. (2009) use the "share of annual arrivals of non-residents in the region who are accommodated in establishments other than hotels, e. g. camping sites or hostels" (p.939) as indicator for openness. For case-based researchers this is unacceptable because tourists do not cause innovation. Keep in mind that variables are distributions across a population, thus, 'level of geographical proximity' is a good variable. Conditions and outcomes are characteristics of cases so case-based researchers will use the condition 'geographically proximate dyads'. Variables are nouns (proximity), conditions are adjectives (proximate) (Ragin and Fiss 2017: 61–63).

Step 3: Calibrating cases. QCA converts raw data into set-membership values. Calibration requires researchers to ask two questions; an ontological one (e. g. what do we mean when we say that firms are geographically proximate) and an epistemological one (how do we know geographically proximate firms when we see them). Cases may be calibrated into crisp sets (in (1) or out of (0) the set) or as fuzzy sets (any value between 0 and 1). In terms of their causal logic and empirical analysis, crisp and fuzzy sets work the same. Fuzzy sets connect better to the empirical intuitions of variable-based researchers, however, for small-N QCA studies (10–20 cases), their added value is negligible. The quality of the calibration depends on researchers' case-based and contextual knowledge and requires transparency on the choices made.

Step 4: Truth-table analysis. Explained in this section.

Step 5: Substantively interpreting empirical findings into causal mechanisms is a matter of 'going back' to the cases. Researchers must answer, e. g. why/how it is possible for geographically proximate AND cognitively proximate dyads to be innovative. Ultimately, it is the plausibility of the substantive interpretation (ontology) rather than the empirical robustness of the set-relationship (i. e., what the truth-table analysis says) (epistemology) that makes a causal argument in QCA.

Table 2: Stylized truth table

Row	conditions				outcome
	cognitive proximity	social proximity	organizational proximity	geographical proximity	innovation
1	1	1	1	1	1
2	1	1	1	0	1
3	1	1	0	1	1
4	1	0	1	1	0
5	1	1	0	0	1
6	1	0	1	0	1
7	1	0	0	1	0
8	1	0	0	0	0
9	0	1	1	1	0
10	0	1	1	0	0
11	0	1	0	1	1
12	0	0	1	1	0
13	0	1	0	0	0
14	0	0	1	0	0
15	0	0	0	1	0
16	0	0	0	0	0

The truth table is analysed on the basis of formal logic. For example, Rows 1 and 2 are characterized by the presence of the outcome (innovation) and the presence of cognitive, social and organizational proximity. They differ in that geographical proximity is present in Row 1 but absent in Row 2. This means that the presence or absence of geographical proximity is logically redundant in these two rows. Consequently, Rows 1 and 2 may be simplified to social proximity AND cognitive proximity AND organizational proximity is sufficient for innovation. All cases that are covered by Rows 1 and 2 are also covered by the simplified configuration, meaning that the truth of Rows 1 and 2 is contained in the simplified configuration. The truth-table analysis is performed by software. This example uses the fsQCA 3.0-software (Ragin and Davey 2017) but other packages, also for R, are available. All QCA software uses the same Quine-McCluskey algorithm and can be downloaded free of cost. The aim of the truth-table analysis is to arrive at a small number of configurations in order to facilitate substantive interpretation. The truth-table algorithm thus simplifies the 16 rows of four conditions each of this truth table to the following three configurations:

- (i) cognitive proximity * social proximity
- (ii) cognitive proximity * organizational proximity * ~geographical proximity
- (iii) social proximity * geographical proximity * ~organizational proximity

Since these are simulated data, there is no case-based or contextual knowledge to substantively interpret the empirical findings. However, the first configuration (i) might suggest that the presence of cognitive proximity AND social proximity allows in-depth learning between dyad partners because the exchange of knowledge and ideas takes place in a context of strong ties. The second configuration (ii) might suggest a form of goal-directed, organized knowledge creation because of the presence of organizational proximity. The absence of geographical proximity suggest that this mechanism only ‘works’ in spatially stretched dyads. However, since organized forms of knowledge creation may also happen between co-located firms, a researcher may perform a separate subset analysis in the software to check whether cognitive proximity AND organizational proximity is also a consistent subset of innovation. If so, the researcher may report this simplified configuration (ii) as the outcome. That is, analytical robustness (plausibil-

ity) is more important than empirical robustness in QCA. The third configuration (iii) might suggest a more serendipitous form of knowledge creation between partners that happen to share a geographical location. The absence of organizational proximity in this configuration underlines the unplanned nature of the causal mechanism.

The truth table in Table 2 is stylized in that the outcome is known for all truth-table rows. In reality, this will (almost) never happen causing real-world data truth tables to have so-called ‘logical remainder rows’. Logical remainder rows may result from researchers not having enough cases to populate all truth-table rows and, more importantly, because some logically possible combinations of conditions (rows) simply do not occur in the real world. Take Row 16 for example. It is characterized by the absence of all proximities. But firms that are not-proximate on any dimension will not form a dyad to collaborate on innovation. Consequently, Row 16 will be empty of cases no matter how many cases are included in the study. This phenomenon is called ‘limited diversity’ and it is a fundamental characteristic of social reality³. Decomposing

social reality into variables blinds researchers for limited diversity but QCA is designed to deal with it (Ragin 2008: 147–149, Ragin and Fiss 2017: 12–13).

For the sake of simplicity; suppose we have observed the following (Table 3):

Table 3: Easy and difficult counterfactuals

Assumptions: the presence of GeoProx, CogProx, SocProx contributes to INNOV

Row	GeoProx	CogProx	SocProx	INNOV	
(a)	1	1	1	?	easy counterfactual
(b)	1	1	0	1	
(c)	1	0	1	1	
(d)	1	0	0	?	difficult counterfactual

INNOV (innovation), GeoProx (geographical proximity), CogProx (cognitive proximity), SocProx (social proximity)

Table 4: Truth table with limited diversity

Row	cognitive proximity	social proximity	organizational proximity	geographical proximity	number of cases	Innovation
1	1	1	1	1	5	1
2	1	1	1	0	–	?
3	1	1	0	1	3	1
4	1	0	1	1	2	0
5	1	1	0	0	3	1
6	1	0	1	0	2	1
7	1	0	0	1	–	?
8	1	0	0	0	–	?
9	0	1	1	1	2	0
10	0	1	1	0	–	?
11	0	1	0	1	3	1
12	0	0	1	1	2	0
13	0	1	0	0	1	0
14	0	0	1	0	2	0
15	0	0	0	1	–	?
16	0	0	0	0	–	?

³ For example; there is no deprived neighbourhood where residents have low incomes, poor housing, are unemployed, have unhealthy lifestyles, but are highly educated.

Based on the available information, configurations (b) and (c) cannot be simplified any further. But if theoretical, contextual and in-depth case-knowledge suggests that it is the presence of cognitive and social proximity rather than their absence that contributes to innovation (the assumptions in Table 3), then a researcher is justified assuming that if a dyad (case) existed for configuration (a), it would be an innovative dyad. That is, configuration (a) is an easy counterfactual. Assuming that (a) is true allows simplifying (b) to geographical proximity AND cognitive proximity (because then the presence or absence of social proximity does not matter for the presence of innovation), and (c) to geographical AND social proximity. A researcher could further assume that, if a case (dyad) existed for configuration (d), it would also be an innovative dyad. This is a difficult counterfactual because in (d) it is the absence, not the

presence, of cognitive and social proximity that would contribute to innovation. If a researcher made this assumption, (c) and (d) could be further simplified to only geographical proximity being sufficient for innovation (Ragin 2008: 161–175). (Note that QCA software automates the process of making assumptions on logical remainder rows.)

Table 4 presents a truth table similar to the one in Table 2 but now with five logical remainder rows. (See Annex 1 for the simulated data.) Analysing the truth table using only the information contained in the 11 ‘populated’ rows produces QCA’s complex solution. Making assumptions on easy counterfactuals only produces the intermediate solution. Making assumptions on difficult counterfactuals also, produces the parsimonious solution. For the truth table in Table 4, these three solutions are as follows (Table 5):

Table 5: Complex, intermediate and parsimonious solutions

complex solution	intermediate solution	parsimonious solution
cognitive * social * organizational	cognitive * social	cognitive * social
cognitive * social * geographical		
cognitive * organizational * ~social * ~geographical	cognitive * organizational * ~geographical	cognitive * ~geographical
social * geographical * ~organizational	social * geographical * ~organizational	geographical * ~organizational

Note that the complex solution is a subset of the intermediate solution, which in turn is a subset of the parsimonious solution. This means that, formal-logically speaking, the solutions do not contradict one another. The configurations in the complex solution positively explain the outcome because they are based on actually observed cases. But the complex solution is, as it says, complex and difficult to interpret. Making assumptions on easy counterfactuals only, i. e. assumptions that are plausible given contextual and in-depth case-knowledge, the intermediate solution is an empirically valid and substantively robust explanation of the outcome that is much easier to interpret. The parsimonious solution identifies configurations that are minimally necessary for sufficiency but that may not substantively explain the outcome (Duşa 2019; Rutten, forthcoming). QCA researchers prefer the intermediate solution because it balances generalizability and interpretability. The intermediate solution is reported in QCA’s solution table (Figure 3).

The solution table reports the configurations identified by the truth table and labels each configuration on the basis of substantively interpreting the nature of the causal mechanism that it reflects. The table shows which conditions are present or absent in each configuration and iden-

tifies which cases are covered by each configuration. As cases are usually covered by multiple configurations (e. g. Area C, Figure 2), the raw coverages may be high but the unique coverages (cases that are covered only by the configuration in question) are usually very low. Consistency shows the extent to which a configuration is a consistent subset of the outcome. In this example, consistencies are 1, meaning that there are no inconsistent cases (i. e. Cell II, Figure 1 would be empty). A high consistency⁴ suggests that the configuration is a good explanation of the outcome. The coverage of configurations is not very relevant. A solution coverage of 1 means that, in this example, all innovative dyads are covered (explained) by at least one of the configurations in the solution. A high solution coverage suggests that the solution is good. Solution consistency is not particularly important.

The most important conclusion is that none of the proximities explain innovation by themselves. Whether a particular proximity, e. g. geographical proximity, matters depends on context. This contextualized notion of causality challenges the causal effects approach, which suggests that, regardless of context, some factors matter (have a

⁴ Consistency should be ≥ 0.8 .

	CONFIGURATION 1 in-depth knowledge creation	CONFIGURATION 2 organized knowledge creation	CONFIGURATION 3 serendipitous knowledge creation
Cognitive proximity	●	●	
Social proximity	●		●
Organizational proximity		●	⊗
Geographical proximity		⊗	●
Cases	01, 02, 03, 04, 05, 06, 07, 08, 11, 12, 13	14, 15	06, 07, 08, 19, 20
Raw coverage	0.6875	0.1250	0.3750
Unique coverage	0.5000	0.1250	0.1875
Consistency	1.0000	1.0000	1.0000
Solution coverage	1.0000		
Solution consistency	1.0000		

● condition present ⊗ condition absent [blank] condition logically redundant

Figure 3

strong effect on the outcome) while others do not. QCA thus strikes a balance between contextualization and generalization. The (intermediate) solution identifies causal mechanisms that are contextualized (case-specific) enough to identify ‘things’ that actually happen, but stylized enough to make them generalizable beyond the cases under investigation (Goertz 2017: 236–239).

Taking stock

The above discussion of QCA is limited and serves two main purposes: 1) to make it clear that variable-based and case-based methods of causal inference follow very different causal logics, and 2) to demonstrate how QCA works. The fact that QCA is itself a multimethod approach that combines (qualitative) substantive interpretation and (quantitative) Boolean algebraic calculation has caused it to come under attack from both sides. Some qualitative researchers have criticized the use of counterfactual analysis (making inferences from cases that are not observed) and the stylized way of reporting results in the solution table. But, generally, qualitative researchers have welcomed QCA (e.g. Beach and Pedersen 2013, Byrne 2009, Kent 2009). Quantitative researchers, emphasizing empirical regulari-

ties, have dismissed QCA because the truth-table analysis is very case sensitive. This is true, but that is exactly why case-based researchers, following a substantive-interpretation logic, like it. The truth-table analysis identifies cross-case patterns. Adding or removing cases qualitatively changes the nature of the (constructed!) population. Changing the calibration qualitatively changes the nature of the cases themselves. Consequently, one would then expect to find different cross-case patterns (Emenegger et al. 2014; Rutten, forthcoming; Skaaning 2011). Others have argued that QCA is not good at dynamics and processes. This is true for the truth-table analysis, not for substantive interpretation. Limitations that cannot be argued away are that QCA can only handle a limited number of conditions and that it is a very labour-intensive method.

The differences between variable-based (epistemological, empirical regularities, predicting outcomes) and case-based (ontological, substantive interpretation, detecting causes) methods identified in this paper are summarized in Table 6. One can look at them and think: we’ll keep it in mind when we analyse the data. But there is rather more to it than that. First, calibrating cases challenges the notion of looking for (co-)variation across populations because not all variation is equally relevant. Second, if one believes that different causal mechanisms may explain innovation

(equifinality), then looking for the effect of any one proximity on innovation is not helpful. The same proximity may play a different role in different causal mechanisms. Third, if one believes that limited diversity matters, then variable-based methods become suspicious because they ignore it. The causal effects approach assumes that variables have an independent effect on the outcome. But if institutional proximity means shared norms and values,

it is difficult to see how dyads can be institutionally proximate without also being socially, organizationally, cognitively or geographically proximate because there has to be a shared context for there to be shared institutions. That is, the notion of limited diversity challenges one of the corner stones of the variable-based approach. Multi-collinearity is not a nuisance that complicates empirical analysis but a defining characteristic of social reality.

Table 6: Variable versus case-based causality

Variable based	Case based
<ul style="list-style-type: none"> – Given populations of diverse cases (samples reflect diversity in population) – Decomposing cases (social reality) into independent and dependent variables – All variation across the population is equally relevant – Effects of causes: everything else being constant, individual causes have an effect on the outcome; effects are ‘law like’ – Probabilistic: causes increase the likelihood of the outcome – Robustness of empirical patterns evidences causality – Formalized with linear algebra (statistics), which identifies simple causality (additive, unifinal, symmetrical) 	<ul style="list-style-type: none"> – Constructed populations of comparable cases (case selection) – Cases (social entities) are considered holistically, they are analytically relevant – Distinction between cases different in kind and different in degree – Causes of effects: configurations of causes explain the occurrence of the outcome; explanations are contextual – Possibilistic: causes make the outcome possible, do not determine it occurs – Empirical patterns are substantively interpreted into causal mechanisms – Formalized with Boolean algebra, which identifies complex causality (configurational, equifinal, asymmetrical)

Fourth, QCA’s distinguishing between empirical patterns (the configurations identified by the truth-table analysis) and causal mechanisms is a major philosophical rift with variable-based methods. It means that QCA dovetails with the critical realist distinction between ontology and epistemology, whereas variable-based methods conflate them (Bhaskar 2008: 36–38, Collier 1994: 76–85, Gorski 2018: 26–27, Lawson 2005: 385). For critical realists, a causal mechanism is real (ontology) independent of our knowledge of it (epistemology). This has two major implications. (i) It means that the outcome does not have to occur for the causal mechanism to be real. Therefore, inconsistent cases do not necessarily defeat causal mechanisms suggested by consistent cases and empirical robustness measures are not decisive for causal inference. (ii) Empirically observed patterns are socially constructed. The tools (methods and theories) used to observe patterns are always biased. They allow researchers to see some things rather than others. Consequently, empirical patterns are not so much observed as constructed by the researcher. This is a position that social constructionist

might embrace (although they are not necessarily into causal analysis) but it puts the emphasis on empirical regularities of variable-based methods in a philosophically uneasy position.

The point of this paper is to demonstrate that different causal logics are indeed very different. Theorizing in terms of effects of variables is something very different than theorizing in terms of causal powers of cases. Since theorizing and data analysis are closely linked, it is vital to understand that, for example, QCA and regression analysis are rather more than different techniques for data analysis. The promise of QCA, and case-based methods in general, is that it opens up exciting new research avenues (viz. complex causality) that are inaccessible to variable-based methods because they are designed to do different things. The ambition of QCA to go beyond causal effects is a challenge to economic geography to look at reality through a case-based rather than a variable-based lens. Given the dominance of variable-based thinking in both theorizing and empirical analysis, QCA has a huge potential in economic geography.

References

- Abbott, A. (1988). Transcending general linear reality, *Sociological Theory*, 6(2): 169–186.
- Abbott, A. (1998). The causal devolution, *Sociological Methods and Research*, 27(2): 148–181.
- Duša, A. (2019). Critical tension: Sufficiency and parsimony in QCA, *Sociological Methods and Research*, published online, pp. 1–25.
- Beach, D. and Pedersen, R. (2016). *Causal case study methods: Foundations and guidelines for comparing, matching, and tracing*, Ann Arbor, MI: University of Michigan Press.
- Bhaskar, R. (2008) (original 1975). *A realist theory of science*, London: Verso.
- Bhaskar, R. (2014) (original 1979). *The possibility of naturalism: A philosophical critique of the contemporary human sciences*, London: Routledge.
- Boschma, R. (2005). Proximity and innovation: A critical assessment, *Regional Studies*, 39(1): 61–74.
- Byrne, D. (2009). Complex realist and configurational approaches to cases: A radical synthesis. In Byrne, D. and Ragin, Ch. (Eds), *Op cit.*, pp. 101–112.
- Byrne, D. and Ragin, Ch. (Eds) (2009). *The Sage Handbook of case-based methods*, London: Sage.
- Collier, A. (1994). *Critical realism: An introduction to Roy Bhaskar's philosophy*, London: Verso.
- Easton, G. (2010). Critical realism in case study research, *Industrial Marketing Management*, 39(1): 118–128.
- Emmenegger, P., Schraff, W. and Walter, A. (2014). QCA, the truth-table analysis and large-N survey data: The benefits of calibration and the importance of robustness tests, *Compass Working Paper 2014–79*, pp. 1–36.
- Gambardella, A., Mariani, M. and Torrì, S. (2009). How provincial is your region? Openness and regional performance in Europe, *Regional Studies*, 43(7): 935–947.
- Gerrits, L. and Verweij, S. (2013). Critical realism as a meta-framework for understanding the relationship between complexity and Qualitative Comparative Analysis, *Journal of Critical Realism*, 12(2): 166–182.
- Goertz, G. (2017). *Multimethod research, causal mechanisms, and case studies: An integrated approach*, Princeton, NJ: Princeton University Press.
- Goertz, G. and Mahoney, J. (2012 a). *A tale of two cultures: Qualitative and quantitative research in the social sciences*, Princeton, NJ: Princeton University Press.
- Goertz, G. and Mahoney, J. (2012 b). Concepts and measurement: Ontology and epistemology, *Social Science Information*, 51(2): 205–216.
- Gorski, Ph. (2018). After positivism: Critical realism and historical sociology. In Rutzou, T. and Steinmetz, G. (Eds). *Critical realism, history, and philosophy in the social sciences*, Bingley, UK: Emerald, pp. 23–46.
- Gorski, Ph. (2013). What is critical realism? And why should you care?, *Contemporary Sociology*, 42(5): 658–670.
- Greckhamer, Th., Furnari, S., Fiss, P. and Aguilera, R. (2018). Studying configurations with Qualitative Comparative Analysis: Best practices in strategy and organization research, *Strategic Organization*, 16(4): 482–495.
- Kent, R. (2009). Case-centred methods and quantitative analysis. In Byrne, D. and Ragin, Ch. (Eds), *Op cit.*, pp. 184–207.
- King, G., Keohane, R. and Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*, Princeton, NJ: Princeton University Press.
- Lagendijk, A., Kuijper, M. and Van der Velde, M. (2020). The conditions for regional collaboration in the Netherlands, *Zeitschrift für Wirtschaftsgeographie*, published online.
- Lawson, T. (2005). Economics and critical realism. In Steinmetz, G. (Ed.). *The politics of method in the human sciences*, Durham, NC: Duke University Press, pp. 366–392.
- Li, P. and Bathelt, H. (2020). Headquarters-subsidiary knowledge strategies at the cluster level, *Global Strategy Journal*, published online.
- Little, D. (2015). Mechanisms and method, *Philosophy of the Social Sciences*, 45(4–5): 462–480.
- Mahoney, J. (2001). Beyond correlational analysis: Recent innovations in theory and method, *Sociological Forum*, 16(3): 575–593.
- Ragin, Ch. (2014) (original 1987). *The comparative method: Moving beyond qualitative and quantitative strategies*, Oakland, CA: University of California Press.
- Ragin, Ch. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*, Chicago, IL: University of Chicago Press.
- Ragin, Ch. (1997). Turning the tables: How case-oriented research challenges variable-oriented research, *Comparative Social Research*, 16(1): 27–42.
- Ragin, Ch. (1992). Introduction: Cases of ‘What is a case?’. In Ragin, Ch. and Becker, H. (Eds), *What is a case? Exploring the foundations of social inquiry*, Cambridge: Cambridge University Press, pp. 1–18.
- Ragin, Ch. and Davey, S. (2017). *fs/QCA [computer programme], Version 3.0*, Irvine, CA: University of California.
- Ragin, Ch. and Fiss, P. (2017). *Intersectional inequality: Race, class, test scores and poverty*, Chicago, IL: University of Chicago Press.
- Rutten, R. (forthcoming). Applying and assessing large-N QCA: Causality and robustness from a critical-realist perspective, *Sociological Methods and Research*, accepted for publication.
- Rutten, R. (2019). Openness values and regional innovation: A set analysis, *Journal of Economic Geography*, 19(6): 1211–1232.
- Rutten, R. (2017). Beyond proximities: The socio-spatial dynamics of knowledge creation, *Progress in Human Geography*, 41(2): 159–177.
- Schneider, C. and Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to Qualitative Comparative Analysis*, Cambridge: Cambridge University Press.
- Skaaning, E. (2011). Assessing the robustness of crisp-set and fuzzy-set QCA results, *Sociological Methods and Research*, 40(2): 391–408.

Annex 1: Simulated data

Cases	cognitive proximity	social proximity	organizational proximity	geographical Proximity	innovation
Dyad 01	1	1	1	1	1
Dyad 02	1	1	1	1	1
Dyad 03	1	1	1	1	1
Dyad 04	1	1	1	1	1
Dyad 05	1	1	1	1	1
Dyad 06	1	1	0	1	1
Dyad 07	1	1	0	1	1
Dyad 08	1	1	0	1	1
Dyad 09	1	0	1	1	0
Dyad 10	1	0	1	1	0
Dyad 11	1	1	0	0	1
Dyad 12	1	1	0	0	1
Dyad 13	1	1	0	0	1
Dyad 14	1	0	1	0	1
Dyad 15	1	0	1	0	1
Dyad 16	0	1	1	1	0
Dyad 17	0	1	1	1	0
Dyad 18	0	1	0	1	1
Dyad 19	0	1	0	1	1
Dyad 20	0	1	0	1	1
Dyad 21	0	0	1	1	0
Dyad 22	0	0	1	1	0
Dyad 23	0	1	0	0	0
Dyad 24	0	0	1	0	0
Dyad 25	0	0	1	0	0