Case-based methods and agent-based modelling: Bridging the divide to leverage their combined strengths

Two leading camps for studying social complexity are case-based methods (CBM) and agent-based modelling (ABM). Despite the potential epistemological links between 'cases' and 'agents,' neither camp has leveraged their combined strengths. A bridge can be built, however, by drawing on Abbott's (1992) insight that "agents are cases doing things", Byrne's (2009) suggestion that "cases are complex systems with agency", and by viewing CBM and ABM within the broader trend towards computational modelling of cases. To demonstrate the utility of this bridge, we describe how CBM can utilise ABM to identify case-based trends; explore the interactions and collective behaviour of cases; and study different scenarios. We also describe how ABM can utilise CBM to identify agent types; construct agent behaviour rules; and link these to outcomes to calibrate and validate model results. To further demonstrate the bridge, we review a public health study that made initial steps in combining CBM and ABM.

Keywords: social complexity, case-based methods; agent-based modelling; qualitative comparative analysis; simulation; social research.

Reflecting on the Potential of ABM and CBM

Given the potential utility of their combined strengths for modelling social complexity, it is our view that the merger of agent-based modelling (ABM) and case-based methods (CBM) has much to offer the social sciences. Despite such potential, researchers have yet to leverage such a combination. Three reasons. First, while ABM is generally focused on simulating social processes for theory testing or applied scenario analysis, CBM focuses on pattern recognition in real data; hence they have developed along different intellectual trajectories (Haynes 2017). Second, ABM requires a basic knowledge of programming, and is often employed by those grounded more squarely in the quantitative tradition; while those using CBM, particularly *qualitative comparative analysis* (QCA), tend to be qualitative researchers (Castellani et al. 2015a; Yang and Gilbert 2008). Third, ABM and CBM have a different approach to modelling, which has sometimes been misconstrued as a difference between a restrictive versus generalist view of complexity – and which has incorrectly led CBM researchers to be somewhat dismissive of ABM and vice versa (Keuschnigg, Lovsjö, and Hedström 2018). This view, however, is misguided, as ABM is a form of *general complexity* (Keuschnigg, Lovsjö, and Hedström 2018). As defined by Morin (2007), generalist complexity "tries to comprehend the relations between the whole and the parts. The knowledge of the parts is not enough, the knowledge of the whole as a whole is not enough…. Thus, the principle of reduction is substituted by a principle that conceives the relation of whole-part mutual implication" (p. 6). Based on this definition, ABM is a type of generalist complexity, as its purpose, as outlined by Gilbert and Troitzsch (2005), is to explore how the microscopic interactions of a set of agents (the parts) lead to emergent forms of complex behaviour (the whole). However, given one its main purposes is to test a theory's capacity to explain the rules governing such complex dynamics (for example, the spread of a disease across a population), it tends to keep things as simple as possible; otherwise one is left unable to tease out a useful model of causality (Barbrook-Johnson et al. 2017).

Still, irrespective of our third point, the differences between ABM and CBM are not sufficient to treat them as methodologically incompatible. In fact, their differences make them useful to each other – hence the purpose of the current paper. As we will show, a methodological bridge can be built between CBM and ABM, mainly by exploring the epistemological links between the concepts of *agency* and *cases*; which allows for several advances in both methods. In particular, CBM researchers can design or use various ABMs to more effectively identify case-based trends across time-space; explore the global dynamics and interactive behaviour of cases; and inspect how different scenarios might impact case-based outcomes. In turn, ABM researchers can use CBM as a complexity-appropriate data framing and analysis approach to more effectively identify and contextualise the complex rules governing different agents' behaviour; pre-identify the potential agent types and trends in a model; and link these types and trends to key outcomes in the model to calibrate and/or validate a model's results (Gilbert & Troitzsch, 2005).

Our paper is organised as follows. We begin with a quick overview of ABM and then CBM. From there we develop a methodological bridge between these two camps. We then outline the advantages of this bridge by reviewing a public health study that, while limited in the success of its merger of CBM and ABM, nonetheless arrived at insights it would otherwise not have achieved (Castellani et al. 2015b). We end by reflecting on future directions for research.

Agent-Based Modelling

Over the last fifty years ABM has developed into a rigorous methodological approach, grounded in a mature academic literature, which enjoys growing appeal inside and outside of academia, including public policy evaluation (Epstein 2006). A Google Scholar search using the phrase "agent-based modelling," for example, returns over 53,300 hits; and Gilbert and Troitzsch's *Simulation for the social scientist* (2005) has over 3,100 citations.

The main strength of ABM is its capacity to act as a virtual laboratory in which modellers can explore the evolving interactions amongst various *agents* (e.g., individuals, households, firms) and their *environments* (e.g., landscape, social network, metropolitan area), relative to some *outcome of concern* (e.g., traffic patterns, housing migration, the spread of a disease) (Gilbert 2008). Compared to other computational modelling approaches (e.g., system dynamics, micro simulation), ABM is most useful, as Johnson explains (2015a), when one or more of the following conditions is true: (1) the effect of interactions and feedback amongst heterogeneous actors is important to the self-organising emergent behaviour of the entire system; (2) spatial dynamics are important in describing the system; (3) path dependence may be an important element in the social system (i.e., past decisions or states affect future decisions or states); and (4) agents can adapt to interventions or changes in the wider system.

Given these strengths, ABMs are typically developed to serve one of three purposes or some combination thereof (Gilbert 2008; Johnson 2015b; Wilensky and Rand 2015). First, they are used for theory development (Barbrook-Johnson et al. 2017), in which a theory is implemented in a model (typically about the behaviour of individuals, households, or firms, and their interaction) and then systematically tested to assess its ability to generate observed outcomes (i.e. generative sufficiency; see Epstein, 2006).

Second, they are used for applied analysis of a real-world issues. In this case, drawing on the results from empirical research (be it qualitative, quantitative, or both) a model is used to simulate potential interventions, counterfactuals, or future scenarios, with results used to inform decision-making (See Gilbert et al. 2018). In other words, vis-à-vis policy, ABM can explore (without real cost) the capacity for various interventions to drive a complex phenomenon of concern in a more effective or useful direction (Barbrook-Johnson et al. 2017).

And, finally, ABMs are used to support engagement with stakeholders. In such instances, the ABM model or development process is used as a highly effective tool for discussion, facilitation or thinking (Gilbert et al. 2018). Said differently, users and modellers can design and run their models together, whilst varying or editing its parameters to explore and discuss their theories or beliefs about agents' behaviour and their environment; or, alternatively, the various interventions they seek to employ and of which we may be interested in comparing.

Overall, then, ABM is a powerful computational modelling tool. And one, in particular, that offers much to CBM in terms of more effectively modelling issues of case-based agency, the interaction amongst cases, and the impact collective dynamics have on macroscopic patterns and trends (Castellani et al. 2015a).

Case-Based Methods

Presently, case-based methods constitute a compendium of techniques. Examples include single-case probabilities, cluster analysis, case-based reasoning, ethnographies, legal case studies, MDSO/MSDO (most different cases, similar outcome/most similar cases, different outcome) and historical case studies (Byrne and Ragin 2009). Despite such differences, the goal of these methods is roughly the same: to study a case or set of cases ideographically – that is, to gain a more holistic understanding of a topic of concern (Ragin and Rihoux 2009). The simplest example is the *case study,* which is an in-depth investigation of a single case. Most approaches, however, engage in some form of case-oriented comparative or case-comparative analyses – the most popular of which is Ragin's *qualitative comparative analysis* (QCA) (Ragin 2014).

The Power and Appeal of QCA

Over the last three decades, QCA has become a well-established and highly regarded method (Ragin 2014). For example, a Google Scholar search using the phrase "qualitative comparative analysis" returns over 16,700 hits; and Ragin's *The comparative method* (2014) has over 8,600 citations.

The purpose of QCA is to engage in a systematic comparison of a small number of *cases* (e.g., political parties across EU countries), using a set of *Boolean variables*, which simplify the *characteristics* of some set of cases (e.g., views on global warming, neoliberalism), in order to enable case-comparison relative to an *outcome of concern* (e.g., differential support of environmental policy). Because of its strong opposition to variable-based statistics and, in turn, its focus on causal complexity, QCA holds wide appeal amongst social scientists. This appeal comes in a variety of forms.

First, QCA works to bridge the quantitative-qualitative divide. As Ragin states: "Most aspects of QCA require familiarity with cases, which in turn demands in-depth knowledge. At the same time, QCA is capable of pinpointing decisive cross-case patterns, the usual domain of quantitative analysis" (2008, p. 1). Still, QCA's focus on variables (to reiterate) is not statistical in its approach. Instead, QCA takes a 'set-theoretic' approach – which means it is not interested in the 'net effect' that some set of variables has on an outcome(s). It is interested in the nuances of how the presence or absence of certain composite combinations of causal conditions (and their complex relationships) link to different sets of outcomes (Ragin 2014). In other words, similar to scale development and principle components analysis, QCA treats variables as complex configurations, which are used to account for key cross-case differences, vis-à-vis some outcome(s) of concern (Ragin 2014).

<<Table 1>>

Second, the techniques of QCA are relatively easy to employ and are visually intuitive. Which also explains, in part, why qualitative-oriented scholars use it (Rihoux and Meur 2009). An excellent example, as shown in Table 1, is what Ragin (2014) calls the *truth table*, which a visual aide for inspecting datasets for cross-case patterns amongst a set of composite variables; all of which can then be reduced to a series of more focused causal statements for different groups of cases. (For more on QCA software, see http://www.compasss.org/software.htm.)

Third, unlike aggregate statistics, QCA regularly creates more than one causal model. Given its set-theoretic approach, QCA seeks to identify distinctive or dissimilar patterns (i.e., groups of cases) and trends across time/space – which is similar to other data mining and classification techniques, such as cluster analysis. And this is very useful because it allows researchers to look for differences between and within groups – which takes us to the last point.

Fourth, QCA's cataloguing of cases into a series of different configurational arrangements is powerful because it allows researchers to explore counterfactual cases and their corresponding outcomes. For example, rather than finding a one-size-fits all model of what an effective school looks like, QCA researchers would look for those poor functioning schools (the counterfactual) that do slightly better than other poor functioning schools (Byrne and Ragin 2009).

Overall, then, CBM is a useful method for data-driven mapping of complex causality across multiple and different groups of cases. And one that offers much more to ABMs than variable-based, linear statistics. For example, as Yang and Gilbert state: "Although in many social sciences there is a radical division between studies based on quantitative (e.g. statistical) and qualitative (e.g. ethnographic) methodologies and their associated epistemological commitments, agent-based simulation fits into neither camp, and should be capable of modelling both quantitative and qualitative data. Nevertheless, most agent-based models (ABMs) are founded on quantitative data" (2008, p. 175).

Building an Epistemological Bridge between ABM and CBM

As stated in the introduction, there is significant potential to leverage the combined strengths of ABM and CBM. However, to do so we need a methodological bridge between the two camps, which we believe can be built by exploring three key epistemological links between the concepts of *agency* and *cases*.

Link 1: Agents Are Cases Doing Things

The first link between ABM and CBM is based on recognising the extent to which the agents in an ABM can be defined as cases doing things. This link comes from Abbott's chapter *What do cases do?* in Ragin and Becker's *What is a case?* (1992). Abbott's argument is rather straightforward. He begins by defining what, for him, constitutes a case – and it is this definition that we follow throughout our study. A case is either an instance of a conceptual class or larger population (1992, p 53). Conceptual classes are social categories or typologies such as those used in intersectionality theory (e.g., economic status, age, nationality, ethnicity, gender, educational level, etc). In such instances, a *case* is a *type*, such as an affluent, younger, professor as compared to a poor, older, lorry driver. Populations, in turn, are sets of things (e.g., small groups, social networks, companies, cities). In these instances, a *case* is a *subset*, for example residents of the Scottish Borders.

In either instance (types or sub-groups), Abbott explains, cases are linked to agency through the concept of social action. In other words, he explains, "by asking what cases do, I am assuming that the case is an agent" (1992, p. 53). For example, one might ask: what are the differences in the smoking and health behaviours of young professors (type) living in the Scottish Borders (subset) versus older, lorry drivers (type) living in Northern England (subset)? And, in terms of QCA's set-theoretical approach, how do the internal complexities of their respective type/subset profiles account for these differences?

Advantages of Link

Abbott's link between cases and agents – which has been at the empirical heart of QCA for the past 25 years (Ragin 2014) – is useful for our epistemological bridge because it demonstrates the two ways that the agents in an ABM are cases. First, in terms of an ABM's conceptual classes, its catalogue of agent types is the same as a list of case types (e.g., for NetLogo users the 'breeds' in a model). And, in terms of an ABM's population, its subgroups (as in the case of geospatial location) are the same as a list of case subsets. The advantage of recognising these similarities is that it allows ABM researchers to make more systematic use of the CBM concept of cases to frame model development, calibration, analysis and the presentation of results.

Link 2: Cases Are Complex Evolving Systems

The second link between ABM and CBM, which extends Abbott's insight, can be built by recognising the extent to which most cases are complex systems and, therefore, in varying degrees agent-based. This link comes from Byrne and Ragin's *The Sage Handbook of Case-Based Methods* (2009), wherein Byrne (Chapter 5) empirically demonstrates that cases are often best modelled as complex evolving systems, given that they are: (1) comprised of a complex causal configuration of variables; (2) grounded in a wider context; (3) dependent, in part, on their initial conditions; (4) path dependent; and (5) irreducible to their constituent set-theoretical formations and therefore emergent. They are also, variously, (6) agent-based, given that few social scientific phenomena, particularly social complexity, are static or without agency.

For Byrne, by saying cases are agent-based he means that cases are best understood and modelled as self-organizing, emergent, dynamic, nonlinear, and (ultimately) interactive. More specifically, he means that cases are often, as in an ABM, decision-making or behaviour-doing actors – which are often also in interaction with one another. Household migration patterns, as we will see in our case study, are a good example. In other instances, however, cases are better modelled as comprised of multiple forms of agency or, alternatively, sets of agents. A good example, which we will also see in our case study, is a community. Before we proceed, however, it needs to be stated up front that, despite Byrne's empirical insight, cases do not always have to be modelled as complex or agent-based, as the aims of a study might differ. Nonetheless, subsequent research by Haynes (2017) and others has strongly supported Byrne's complex systems view of cases (Castellani et al. 2015a, 2015b; Williams and Dyer, 2017).

Advantages of Link

In terms of CBM, Byrne's complex systems view is useful because it challenges researchers to give more attention to the various ways that their study and its composite variables are agent-based; that is, how cases engage in some form of social action or behaviour – which few QCA studies, for example, explore. In turn, it also challenges CBM researchers to think about how cases interact, how these interactions impact their respective trajectories, and what are the emergent macroscopic consequences of these various interactions, or more generally, collective behaviour. Again, these are forms of analysis that very few QCA studies do. As such, as Haynes has pointed out (2017), thinking about case-based dynamics is a major advance on CBM and, more specifically, QCA method.

Link 3: ABM and CBM as Computational Modelling

The third link between ABM and CBM can be built by recognising how both methods are part of the larger *case-based modelling trend* in computational methods. Before we proceed, however, a caveat is necessary. Unlike the previous two links, the third is not specific to QCA and ABM. Instead, it focuses on connecting ABM to recent advances in computational modelling, which are variously case-based. From this perspective, a typical row vector c_i in a computational model, mathematically speaking, is comparable to a QCA case and its profile. In turn, a database D consisting of row vectors $c_i = (x_{i1}, x_{i2} \dots, x_{ik})$ – even if calibrated using Boolean algebra – is roughly similar to a QCA *truth table*.

Following Witten, et al. (2016), examples of the latest trends in computational modelling include *data mining* (e.g., Bayesian statistics, cluster analysis), *social network analysis* (agent-network theory, complex networks), *data visualisation* (e.g., computer graphics, visual complexity), *machine intelligence* (e.g., genetic algorithms, artificial neural nets), *dynamical systems theory* (e.g., continuous dynamical systems, synergetics), and *geospatial models* (e.g., gravity models, spatial analysis). And all of these methods (albeit to varying degrees) can be counted as an improvement on conventional statistics, mainly because they avoid variable-focused and aggregate-based one-size-fits-all solutions. In other words, they are better at modelling complex causality because (similar to QCA) they are case-based. For example, by focusing on MRI images (as cases), neural nets can identify tumour or disease types and their change over time; genetic algorithms, in turn, can identify reliable trends in stocks (cases) for strong investment opportunities; and, by treating storms or automobiles as cases, differential equations modelling can detect subtle changes in weather or traffic patterns (Witten, et al. 2016).

Advantages of Link

First, the utility of this link is that it widens the definition of case-based methods, in particular QCA, to include the techniques of computational modelling. For example, the public health study that we explore below, while case-based, did not use QCA; instead, it used a combination of k-means cluster analysis and machine intelligence (Castellani et al. 2015b). As shown in Figure 1, it also replaced the truth table with what is known as a u-matrix (topographical neural net). While we cannot delve into the details, a u-matrix is a visual tool for highly sophisticated cross-case comparisons. For example, in this study, it shows the 20 communities in the public health study and their respective cluster membership, as well as their conceptual position relative to every other case and cluster.

<<Figure 1>>

Second, as others have likewise been doing (e.g., Gilbert et al. 2018; Keuschnigg, Lovsjö, and Hedström 2018), this connection allows us to further link ABM with the latest advances in computational modelling, particularly longitudinal methods. Unlike QCA, most computational modelling methods regularly focus on how cases, in the form of trends, evolve across time/space (Han, Pei and Kamber 2011). This improvement in modelling cases longitudinally is key, as it allows us to make an important advance on the field.

To do so, we draw on the work of Rajaram and Castellani (2012, 2015), which makes the connection between the mathematical formalisation of a case as a *row vector* and the mathematical formalisation of a case as a *vector with magnitude and direction*. The first formalisation is familiar to most social scientists, as it is the ‘case’ in a typical statistical database, as defined in matrix algebraic terms and as regularly used in QCA as well. The second formalism, which comes from calculus and physics, is more familiar to simulation scientists and, more specifically ABM, as it focuses on how 'cases,' individually and in terms of their collective dynamics, move across time/space.

Based on Rajaram and Castellani's mathematical connection (2012, 2015), we can extend this idea to relate the cases in a typical quantitative database (e.g., truth table, for example) with their corresponding collective dynamics (particularly geospatial) in an ABM. However, because the mathematics involved in this link are rather detailed, and because Rajaram and Castellani (2012, 2015) have already provide such a proof, we refer readers to those papers, skipping directly to the advantages gained from doing so. The first is that it highlights ABM as form of computational modelling for agent-based interactions and collective dynamics and their emergent macroscopic outcomes (See, for example, Castellani et al 2015a). Second, it indirectly points to the potential of ABMs to be used as clustering techniques – albeit in certain instances and not always – given that one of the activities of designing an ABM, or alternatively making sense of its output, is to group agents into a set of meaningful types, based on different rule configurations and outcomes.

The advantages of linking CBM and ABM

Now that we have a basic sense of ABM and CBM, as well as the methodological bridge that can be built to connect them, it is time to list the advantages that come from such a merger. However, rather than simply provide a summary list, it seems more useful to first review (albeit quickly) a case study where these methods were somewhat effectively combined, which we can then use to better argue our list. We do note however, before proceeding, that the public health study's merger of ABM and CBM was an early attempt, and therefore, at best, a proof-of-concept, with the challenge for additional research to more rigorously test how to more effectively leverage the combined power of these methods.

Case Study

As with most attempts at methodological advance, the study we review here – *Place and health as complex systems: A case study and empirical test* (Castellani et al. 2015b) – was the outgrowth of a research challenge. They asked: to what extent is it useful to conceptualise and model public health (as well as the wider socio-ecological context in which it is situated) in complex systems terms? To explore this challenge, Castellani et al. (2015b) studied the health and wellbeing of twenty communities in a Midwest county in the United States. The substantive challenge was to understand, in particular, why a handful of the poorest urban communities remained caught in a health poverty trap over a ten-year time-period, despite significant public health investment?

To answer this question, the study, which employed a mixed-methods toolkit, turned first to the tools of CBM, in particular, as noted earlier, k-means cluster analysis and machine intelligence, which are both methods of classifying cases into different groups, based on differences in their respective profile of factors (i.e. their k-dimensional vectors) – which, in the case of the current study was a combination of public health and socioeconomic factors – and then tracking their trends (i.e., evolving dynamics and change) across time (for a detailed justification of its methodological approach, see Castellani et al. 2015b).

The results were not entirely unexpected: overall seven clusters were identified. Of these seven, the two clusters with the worst health outcomes were poor, urban communities with a significant proportion of minorities, teenage pregnancies, crime, few home owners, and a large population of living-alone elderly, as well as poor educational outcomes and limited access to healthcare and prevention. In turn, the healthiest communities, which were all in the outer suburbs of the county, were doing very well across all of these factors.

However, because Castellani et al. (2015b) used CBM to search for different trends – rather than linear modelling, which would have explored variables rather than cases, and, in turn, would have searched for one aggregate (bell shaped curve) trend across all 20 communities – they hit on something unexpected. They noticed that whilst the poorest communities did not change over the ten-year period of our study, they did make some progress in job growth, preventative services, etc. However, it seemed that no matter how well they did, the affluent clusters always out-developed them. They also noticed that, over time, the populations in the poorest clusters decreased, whilst the suburban affluent clusters gained in population.

In other words – dropping down a level from the communities as cases to the households within them – it seemed that if a poor household improved socioeconomically it moved to a more affluent community; in turn, if a middle-class household did well it likewise moved to a richer community; and, in turn, those with the highest income levels continued to sequester themselves into smaller and smaller suburban clusters of wealth and privilege – a phenomena known as *suburban sprawl*. And it was the movement of these households (as cases), which seemed to negatively impacted the larger trends in the communities themselves, particularly in terms of the variables normally examined by public health researchers, as outlined above (e.g., poor schooling, poverty, etc).

The challenge, however, was that using only the tools of CBM and its community-level data, Castellani et al. (2015) had no way to test these unexpected insights into the potential role of suburban sprawl, relative to the normal set of public health factors. As such, they turned to the tools of ABM to develop the model they called *Summit-Sim* (i.e., the county they studied is called Summit County, Ohio), which was a basic variant on the famous Schelling model of segregation. Let us explain.

<<Figure 2>>

The purpose of Summit-Sim was to see if the out-migration of upwardly mobile poor, middle-class and rich households (the communities, as cases, turned into micro-level agents) helped to create the macroscopic phenomena they saw in these data, including the poverty traps in which the poorest communities in Summit county were caught. It worked as follows: typical to the United States, rich agents seek to create concentrated suburban neighbourhoods of wealth by moving away from everyone else; meanwhile, middle-class agents seek to live near the rich; and, in turn, poor agents seek to be near the middle-class. Everyone, however, cannot move so easily, given differences in socioeconomic status and wealth; also, the degree to which agents preferred to be around others could be varied, as in Schelling's model, going from mild to severe.

While we cannot explore the details here, Summit-Sim (albeit in simplistic terms) reasonably supported Castellani et al.'s (2015b) hypothesis about the negative impact of sprawl. They found that the micro-level out-migration behaviours of households (their cases) – broken down into three case types of poor (triangle) to middle (star) to rich (square) – did create the same suburban sprawl they saw in their data at the community level, including the creation of secluded communities of affluence (Circle A, Figure 3), a suburban spread of middle-class agents across the map, and (Circle B, Figure 3) health poverty traps comprised almost entirely of poor agents.

Equally important, because their model acted as a virtual lab in which they could explore different scenarios, they also found that, if suburban sprawl was more effectively regulated, the segregation amongst rich, middle and poor agents was less severe, including the dissolution of community-level health poverty traps. Which suggested that one possible policy-based measure for improving poor communities (as in the case of poor schooling, housing and employment instability, and so forth) is to control sprawl.

<<Figure 3>>

As discussed at the end of their study, as a function of combining CBM and ABM – which allowed them to study the interaction between cases as agents at the household level; and to think of communities (i.e., cases) as complex systems comprised of a set of agents – Castellani et al. (2015b) gained a level of insight they would not have otherwise achieved. Still, while the insights gained were significant, the ABM used by Castellani et al. (2015) did not include, for example, any sort of community-level socio-economic constraints; nor did it force the households in Summit-Sim into the same communities (subsets) at the initial stage of the model. Nor did their model simulate how the behaviour of households (its primary cases) impacted how the communities in Summit County, as cases, changed socioeconomically across time. Nonetheless, as an initial proof-of-concept, Castellani et al. (2015b) does suggest there is a real potential for the leveraging of the combined strengths of CBM and ABM, which we will seek to quickly list now, starting first with the advantages for CBM.

Advantages for CBM

Overall, as our case study hopefully suggests, for CBM scholars the main advantage of combining their methods with ABM is that they can more effectively study the behaviours and interactions of cases; the impact these social inter-actions have on their respective trajectories and trends; and, in turn, the larger emergent macroscopic systems of which they are a part. Such an advance is significant, particularly for QCA, because other than a small set of specific methods, such as dynamic pattern synthesis (Haynes 2017) and case-based density modelling (Rajaram and Castellani 2012), most CBMs are not designed to study multiple longitudinal trends across time, or they do not do so as effectively as ABMs.

We acknowledge, however, that in many instances a CBM study may not be interested in what its cases are doing. Instead, it might simply be focused on identifying key patterns and multiple subgrouping of causal complexity. At other times, however, CBM scholars may want to know what their cases are actually doing. And, even further, scholars may want to know what these cases are doing in interaction with other cases. While in other instances CBM scholars may be interested in exploring the agency of cases at multiple levels, as in the study of collective dynamics and macroscopic trends demonstrated in our case study.

As such, during the study design and data collection processes, thought should be given to if, when, and how the variables in a case profile or, more specifically a QCA *truth table* (even if expressed in Boolean algebra) are manifestations of social interaction or agent-based behaviour of some type. And, if warranted, researchers can then move from these results, as demonstrated by Castellani et al. (2015b), to think through what questions they would like to answer and therefore design their ABM to explore. It is at this point that we recommend reaching out to the ABM community, as there may be models that presently exist that CBM researchers could use or adapt, or alternatively new models that they need help developing. We would recommend beginning such a 'reach out' with dedicated journal such as the *Journal of Artificial Societies and Social*

Simulation, or relevant learned societies such as the *European Social Simulation Association*, or the *Computational Social Science Society of the Americas*.

The other major advance that ABM provides for CBM is that, once a model has been developed, it provides the capacity to further explore counterfactuals and to inspect how different scenarios or interventions might impact case-based outcomes or drive a study in a different or more desired direction, as in the case of public policy or social services. For example, in Castellani et al (2105b), their ABM was not limited to the constraints of their CBM empirical data. Instead, they were able to explore a variety of anti-sprawl scenarios and counterfactuals conditions (using a series of sensitivity analyses) to see if there was a way to effectively reduce the negative impact that the outmigration of affluent household (cases as agents) had on poor households in the model.

Advantages for ABM

The main advantage CBM provides ABM is the capacity to engage in a more sophisticated preliminary investigation of the causal complexity it seeks to simulate. In other words, to repeat an earlier point, CBM allows ABM researchers to more explicitly and formally connect together – under a common goal of embracing rather than reducing complexity – CBMs that cluster or catalogue cases and their complex causality with their ABMs, which study the collective dynamics of these cases (as agents) in complex systems terms across time/space. Such an advance is significant because, beyond the collection of qualitative or historical data, current convention in ABM relies heavily on conventional variable-based statistics for use in the model-building phase, specifically the design and validation of micro-level agent rules (Yang and Gilbert 2008). These traditional approaches provide analyses that contradict the complexity-based epistemology of ABM. By making use of CBM analyses in the model design phase, ABM researchers will no longer have to take part in this epistemological cognitive dissonance.

In terms of the specifics of model design, using or conducting a CBM analysis has the following advantages. First, it would provide ABM researchers further information from which to identify the different agent types for their model. In the case of Castellani et al. (2015b), for example, the results of their CBM inquiry allowed them to identify and validate the use of three key agent types: rich, middle and poor households.

Second, it would allow ABM researchers to more effectively calibrate their models (e.g., choose the best micro-level agent or model designs and parameter values that make the model produce plausible results) and create the rules and conditions that govern the behaviour of different agents. For example, in the case of Castellani et al. (2015b), they were able to realise that the key rules revolved around rich agents trying to escape into suburban neighbourhoods of privilege and position, while chased closely behind by middle agents, who were being pursued by the poor but upwardly mobile households. They were also able to write these rules as a continuum from very aggressive outmigration to restricted outmigration, which allowed them to test varying levels of segregation.

More abstractly, the outputs of CBM analysis – in which casual complexity is described more fully for a particular setting – provide modellers a richer picture of the factors (i.e. different configurations of factors associated with an outcome) that are important to model or include in their micro-level agent rules. In the case of Castellani et al. (2015), for example, this picture included larger deindustrialisation trends in the Midwest and the turn by the middle and professional classes to a life in the suburbs.

Finally, using CBM allows ABMs to link their different agent types and their corresponding trends to key outcomes to empirically validate the complex emergent results of an ABM – which are often difficult to narrate and interpret, or are compared uncritically to traditionally aggregated data (i.e. using averages). For example, Castellani et al. (2015b) were able to take the results of their suburban sprawl model and compare its results with actual geospatial data of out-migration trends (broken down in the same way as their case groups) in the county they studied, which they found did reasonably support the community-level insights of their model. However, given the limitations and lack of available data, they were not able to empirically validate the model's insight that a more restricted approach to suburban sprawl would dissolve the community-level health poverty traps they found in their data.

More abstractly, there are two key ways CBM analysis could be used to aid in model validation. First, micro-level outcomes could be validated using the findings of CBM analysis; that is, patterns that are observed in real data using CBM could be looked for in model behaviour. Second, real-world data used in model calibration and validation could be aggregated or re-framed in case-based forms, or indeed data could be collected in case-based forms, to allow the model to validate against more appropriate benchmarks (i.e. rather than against population averages which do not capture non-normal distributions).

Conclusion

While the current study identified some key ways to link CBM and ABM and the consequent advantages that can come from doing so, further research is necessary to develop the ABM/CBM link. In particular, we believe it would be fruitful to further develop and operationalise some of the conceptual links we have detailed above. For example, it would be useful to examine how the usage of social action and interaction variables in a QCA truth table might lead to more usable and validated design of agent rules in an ABM; or, in turn, how ABMs could corroborate the different configurational arrangements across time found in a discrete QCA study. Further, it would be valuable to explore how a hybrid CBM/ABM method (or at least a more formal protocol for how they can complement one another) might be developed that exists somewhere in the middle of these two methods. Beyond these specific avenues for which we see potential progress, we hope this paper brings these two methodological communities closer together and facilitates the combination of the conceptual and analytical tools of each in whichever forms individuals or groups of researchers see fit.

REFERENCES

Abbott, A. (1992). What do cases do? Some notes on activity in sociological analysis. *What is a Case*, 53-82. In Ragin, C. C., & Becker, H. S. (Eds.). *What is a case?: exploring the foundations of social inquiry*. Cambridge university press.

Barbrook-Johnson P, Badham J, Gilbert G (2017) Uses of agent-based modeling for health communication: The TELL ME case study, Health Communication 32 (8) pp. 939-944

Burrows, R., & Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big data & society*, *1*(1),

Byrne, D., & Callaghan, G. (2013). *Complexity theory and the social sciences: The state of the art*. Routledge.

Byrne, D., & Ragin, C. C. (2009). *The Sage handbook of case-based methods*. Sage Publications.

Capra, F., & Luisi, P. L. (2014). *The systems view of life: A unifying vision*. Cambridge University Press.

Castellani, B., & Hafferty, F. W. (2009). *Sociology and complexity science: a new field of inquiry*. Springer.

Castellani, B., Rajaram, R., Gunn, J., & Griffiths, F. (2015a). Cases, clusters, densities: Modeling the nonlinear dynamics of complex health trajectories. *Complexity*, *21*(S1), 160-180.

Castellani, B., Rajaram, R., Buckwalter, J. G., Ball, M., & Hafferty, F. (2015b). *Place and health as complex systems: A case study and empirical test*. Springer.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems*. Routledge.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Gilbert G, Ahrweiler P, Barbrook-Johnson P, Narasimhan K, Wilkinson H. (2018) Computational Modelling of Public Policy: Reflections on Practice, Journal of Artificial Societies and Social Simulation 21 (1) pp. 1-14

Gilbert, N. G. (2008). Agent-based Models. Sage Publications Ltd.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Johnson, P. G. (2015a). *The SWAP model: Policy and theory applications for agent-based modelling of soil and water conservation adoption* (Doctoral dissertation, University of Surrey (United Kingdom)).

Johnson, P. (2015b) Agent-Based Models as “Interested Amateurs”, Land 4 (2) pp. 281-299.

Haynes, P. (2017). *Social Synthesis: Finding Dynamic Patterns in Complex Social Systems*. Routledge.

Keuschnigg, M., Lovsjö, N., & Hedström, P. (2018). Analytical sociology and computational social science. *Journal of Computational Social Science*, *1*(1), 3-14.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining*. John Wiley & Sons.

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Morin, E. (2007). Restricted complexity, general complexity. In *Worldviews, science and us: Philosophy and complexity* (pp. 5-29).

Ragin, C. C. (2014). *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press.

Ragin, C. C., & Becker, H. S. (Eds.). (1992). *What is a case?: exploring the foundations of social inquiry*. Cambridge university press.

Ragin, C., & Rihoux, B. (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques.* Sage.

Ragin, C. C., & Fiss, P. C. (2016). Intersectional Inequality. *Race, Class, Test Scores, and Poverty*. University of Chicago Press.

Rajaram, R., & Castellani, B. (2015). The utility of nonequilibrium statistical mechanics, specifically transport theory, for modeling cohort data. *Complexity*, *20*(4), 45-57.

Rajaram, R., & Castellani, B. (2012). Modeling complex systems macroscopically: Case/agent-based modeling, synergetics, and the continuity equation. *Complexity*, *18*(2), 8-17.

Rihoux, B. & Meur, G. (2009). Crisp-set qualitative comparative analysis (csqca). In Rihoux, B., & Ragin, C. C. *Configurational comparative methods: Qualitative comparative analysis (QCA)*. Thousand Oaks, CA: SAGE

Williams, M., & Dyer, W. (2017). Complex realism in social research. *Methodological Innovations*, *10*(2), 2059799116683564.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Yang, L., & Gilbert, N. (2008). Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in complex systems*, *11*(02), 175-185.

FIGURE AND TABLES

Table 1. An Example of a Truth Table with 3 Cases, 2 Variables and 1 Outcome

Table 1. QCA Truth Table with 3 Cases, 2 Variables and 1 Outcome

Case	Variable 1 Economic Growth	Variable 2 Healthcare Access	Positive Health Outcome Community-Level Mortality
1	0 (No)	0 (No)	0 (No)
2	1 (Yes)	1 (Yes)	1 (Yes)
3	1 (Yes)	1 (Yes)	1 (Yes)

Figure 1: Example of a Neural Net U-Matrix, as created for a public health study of a county and its 20 communities.

Neural Net U-Matrix as alternative to the QCA truth table

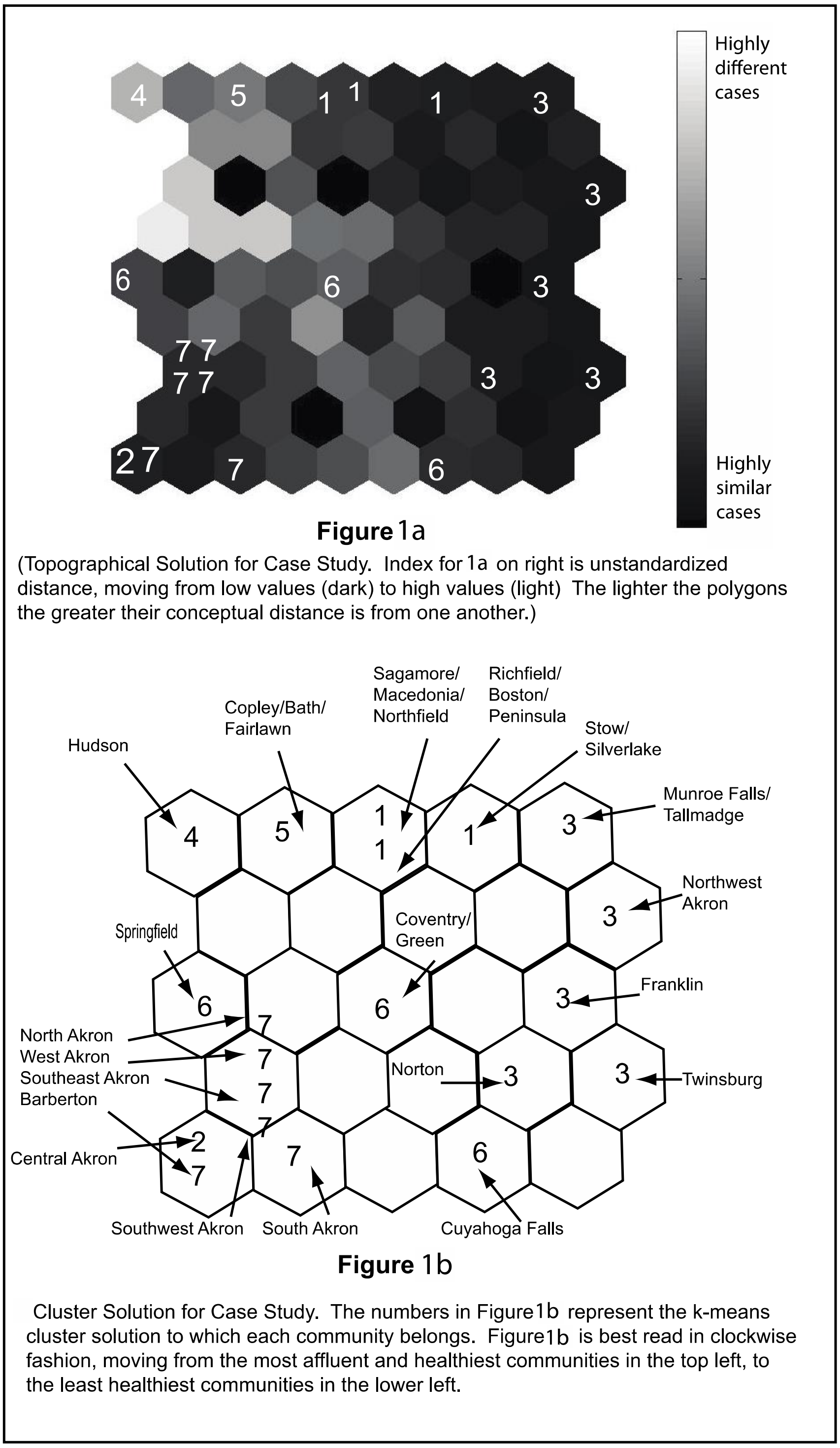

Figure 1a

(Topographical Solution for Case Study. Index for 1a on right is unstandardized distance, moving from low values (dark) to high values (light) The lighter the polygons the greater their conceptual distance is from one another.)

Figure 1b

Cluster Solution for Case Study. The numbers in Figure1b represent the k-means cluster solution to which each community belongs. Figure1b is best read in clockwise fashion, moving from the most affluent and healthiest communities in the top left, to the least healthiest communities in the lower left.

Figure 2: Snapshot of the ABM Model to Explore Suburban Sprawl and Poverty Traps

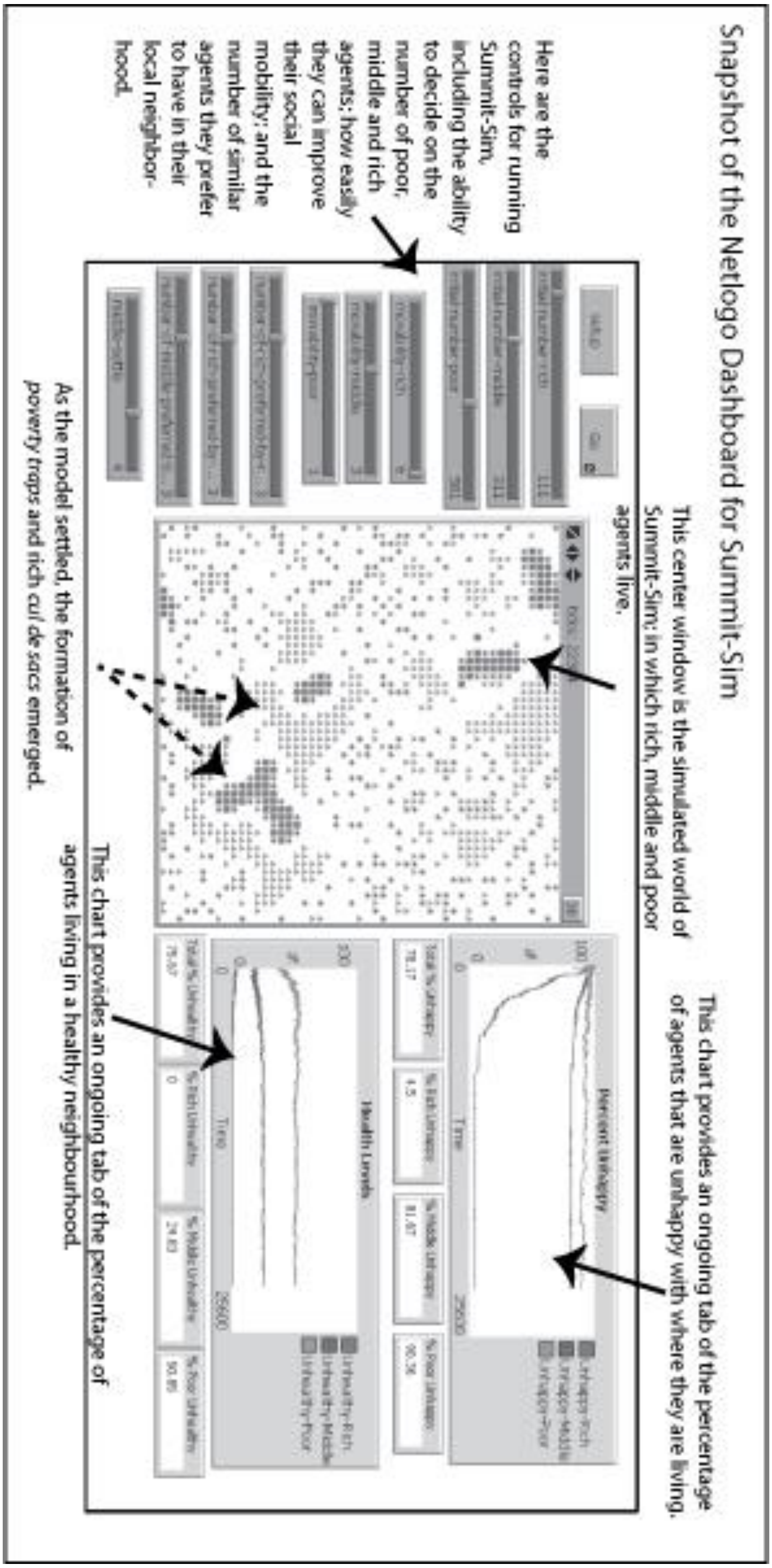

Figure 3: Snapshot of ABM Model Solution Demonstrating the Presence of Poverty Traps as a Function of Suburban Sprawl.