A ‘Theory of the Middle Range’ to Support Food Security and Circular Economy Value Chain Scenario Analysis*

Gary Polhill[1] [0000−0002−8596−0590],
Benjamin J. J. McCormick[2] [0000−0002−8060−0502],
Nick Roxburgh[1] [0000−0002−7821−1831], Samuel Assefa[1] [0000−0001−9557−3810], and
Keith Matthews[1] [0000−0001−8472−8872]

[1] Information and Computational Sciences Department, The James Hutton Institute, Aberdeen AB15 8QH, Scotland, UK
{gary.polhill,nick.roxburgh,samuel.heban,keith.matthews}@hutton.ac.uk
https://www.hutton.ac.uk/social-simulation/
[2] The Rowett Institute, University of Aberdeen, Aberdeen AB24 3FX, Scotland, UK
benjamin.mccormick@abdn.ac.uk
https://www.abdn.ac.uk/rowett/

Abstract. A common critique of agent-based models is that their relationship with theory is tenuous. However, arguments have been made showing both that agent-based models can draw on theories, and that agent-based models can be treated as though they are theories. At the same time, empirical agent-based models, increasingly used for policy evaluation and scenario analysis, tend to be fitted to specific cases rather than having any sense of generality to their application. This lends weight to the criticisms of the relationship between agent-based models and theory. To address this issue in a project in which we are tasked with creating an agent-based model to evaluate scenarios in ‘farm-to-fork’ value chains, with potential applications to more general circular economy case studies, we have endeavoured to design a ‘theory of the middle range’ (à la Merton) that we believe to have the necessary capability, together with a formal implementation in the form of an agent-based model. We describe the model here, giving critical attention to its proposition as a ‘theory’ (of any kind) and potential generality, with a view to furthering the debate about the relationships between empirical agent-based models and social/scientific theories.

Keywords: Value Chains · Ontologies · Middle-Range Theories · Agent-Based Models

* This work was supported by the Scottish Government Rural and Environment Science and Analytical Services Division (project references JHI-C5-1, JHI-C3-1, JHI-C4-1) and the Research Council of Norway (project number 294777).

1 Introduction

This paper is concerned with theory development, which Lorscheid et al. [8] have called for more of in the light of criticisms such as those of O'Sullivan et al. [12] that agent-based models do not lead to much of it. To this end, Lorscheid and colleagues organized a series of workshops on theory development with agent-based models, and a special issue of *Environmental Modelling & Software* is in preparation, though regrettably not at a stage at which articles submitted to it can be cited here.

Some of the earliest articles in *JASSS* are exercised by the relationship between agent-based modelling and game theory (e.g. [10]), which entailed many of the arguments subsequently rehearsed by Waldherr & Wijermans [14] and revisited in making the case for theory developments by Lorscheid et al. [8]. In the context of that earlier conversation around game theory, however, Balzer et al. [2] make the assertion (para. 3.16) that "every simulation study has the same status as a scientific theory in its state of being proposed for the first time." This point about initial proposition pertains to the question of whether there is a "group of practitioners," which Balzer et al. [2, para. 3.5] argue is a necessary condition for a scientific theory to have such status.

The arguments Balzer et al. make that there is an equivalence between a simulation study and a scientific theory suggest that perhaps one of the problems in the seemingly interminable discussion about agent-based models and theory is that we do not treat our models as such. Should that be the case, then a potential remedy is for those composing agent-based models to posit them as theories, and to proceed with their design and implementation with that high aim in mind. It need hardly be said that there is considerable debate in the philosophy of science on what, exactly, a scientific theory is (see [11]). We therefore need to be clear that we expect the theory to be applicable to empirical case studies, rather than being simply a useful tool to help us think about the world, but not applicable to anything specific in it.

The goal of the article is to report on the experience of taking on the above challenge: developing an agent-based model with the specific intent that it constitutes a theory that can be applied to empirical case studies. The context is a project in which we are interested in using agent-based models to explore scenarios in which policy intervenes in business domains with a view to achieving its aims. The two prime areas of interest are in supporting food security and the circular economy. Gilbert et al. [6] have observed that models for exploring potential policy interventions need not be empirical; and indeed they are not always of the greatest value to policy analysts. In the context of these areas of interest, the prime concern is about 'uptake' of measures, a matter that is sensitive to the specific contexts that the agents in the model (who represent businesses) find themselves. Since we have access to data providing information on those contexts, we might as well represent the empirical situation explicitly. With respect to theory development, since we have taken the view of scientific theories that they should apply to the empirical world, the ambition in positing the model as a theory puts us in a position in which we face the greatest discomfort when the empirical world doesn't quite fit.

On a related note, Boero and Squazzoni's [3] 'spectrum' of agent-based modelling types includes the concept of the 'typification': models that apply to a *class* of empirical case studies rather than to a specific case study. This is all very well from a scientific theory point of view, so long as that class can be instantiated to various specific empirical case studies. Given such a situation, we might regard such 'typifications' as "theories of the middle range" as famously postulated by Merton [9]. Middle-range theories are still theories, but have more constrained generality than would otherwise be expected of a scientific theory. Merton suggested that the more modest aim of middle-range theories could be a way forward for theorization in the social sciences.

The rest of the paper introduces the middle-range theory 'STRAVVS' (Simulating Transacting Rural-Area Ventures in Value Systems), explaining the appeal of the initial concept for modelling material flows in networks of business-to-business exchange. We then document how the ontology of STRAVVS had to be updated, both for the purposes of implementing a model and to make the initially appealing theory more applicable to a specific case study. As may be appreciated, the latter in particular led to complication rather than simplification. We close by discussing what we have learned from the intention at the outset to construct the model as a middle-range theory, and implications for the relationships between agent-based models, theories and the empirical social world.

2 STRAVVS version 0: A middle-range theory of value chains

The central intuition behind the conceptualization of STRAVVS was the idea that businesses (*Processors*) participate in networks of exchange, implementing *Processes* that transform *Ingredients* into *Makes* using *Catalysts* (Figure 1). *Ingredients*, *Makes* and *Catalysts* are all abstractions of *Products*

For example, a cattle farmer specializing in bringing on beef calves uses a tractor and a cattle shed, consuming fertilizer, grass seed, red diesel and specialist feeds to convert beef calves into beef cattle that are bought by a cattle farmer specializing in finishing off. Co-products include slurry, nitrogen dioxide emissions from fertilizer application, methane emissions from ruminant digestion, and any habitat afforded to wild species from these activities. This network forms a largely acyclic spatially-embedded directed graph with root nodes in natural resources and ports, and leaf nodes at consumers. As an economy becomes more circular, the graph becomes more cyclic, and there is less distinction between *Processors* who run the *Processes* and *Consumers* who consume the end products, as the latter have co-products from their consumption that other *Processors* use. Government interventions in that network, in the form of grants, taxes and regulations, are the means by which its policies are realized.

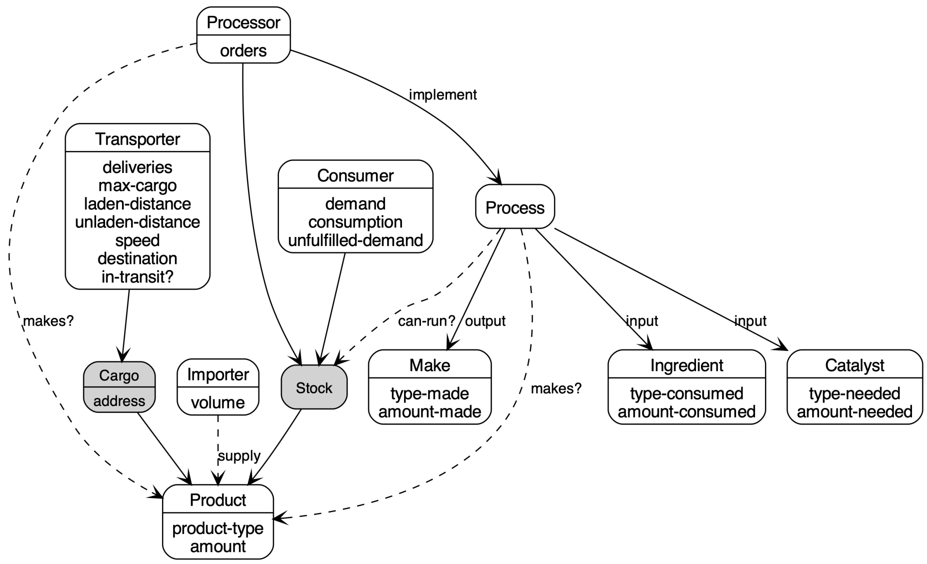

Fig. 1. First version of the STRAVVS ontology. Solid lines are explicit relations (i.e. those implemented as `un/directed-link-breeds`), with 'reified' relations (typically those with `-own` attributes) shown as grey boxes. Dashed lines are implicit relations (implemented as `-own` attributes or using reporters).

The insights of Jarvis et al.'s [7] work on Resource Acquisition, Distribution and End-use networks are that the movement of resources are critical to the functioning of economies, and itself is responsible for a significant amount of consumption. We added *Transporters* to the ontology to reflect this. Finally, to make a functioning network of exchange, we need an *Importer* that brings in *Ingredients* and *Catalysts* not made by a *Process* implemented by any *Processor*.

The theory is not specific about how (or even whether) *Processors* choose the *Processes* they will implement, or the other businesses from which they will source their *inputs*, or for which they will make their *outputs*. These are determined by the specific case studies to which the theory is applied. While the fact that the agents mostly represent businesses means that we might assume a profit maximization strategy, there are other considerations businesses might make, for example: the provenance of the *inputs* might matter, or businesses might value loyalty, or reliability and timeliness of supply. Decisions might be based on heuristics, satisficing rather than optimizing, memories of previous transactions, experiences with running particular *Processes* or trying to sell *outputs*. In some cases, where there is sufficient data to support it, algorithms for decisions might be developed using machine learning; decisions could also be based on formalizing qualitative evidence from documents or interviews.

To test the theory's consistency, we implemented it as an agent-based model. Figure 2 shows a simple spatial network featuring an *Importer* on the left-hand

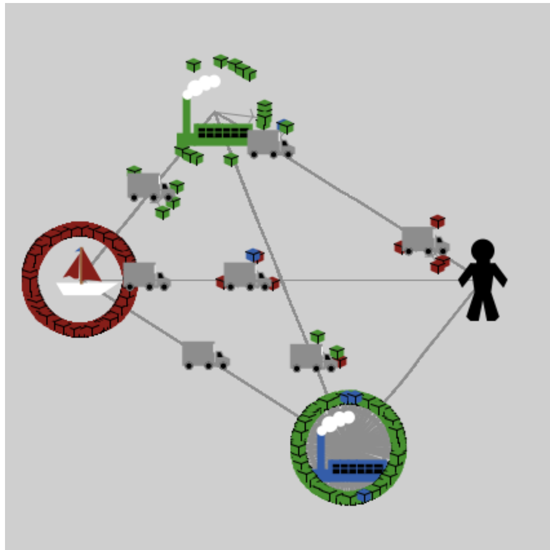

Fig. 2. Screenshot from a simple NetLogo implementation of the STRAVVS theory showing two *Processors* (factory icons) using *Ingredients* (box icons) from an *Importer* (boat icon) to fulfil the demand of one *Consumer* (person icon) via *Transporters* (lorry icons).

side importing *Ingredient* `"A"`, two *Processors* in the middle (one above and slightly to the left of the other), and a *Consumer* on the right-hand side who needs *Product* `"C"`. Two *Processes* are available. The first, implemented by the upper *Processor* takes three units of `"A"` and produces one unit of `"B"`. The second, implemented by the lower *Processor* takes one unit of `"A"`, two of `"B"`, and produces one unit of `"C"`. The *Products* and *Ingredients* are shown in the figure as small boxes, either as the *Stock* of a *Processor*, or as the *Cargo* of a *Transporter*, of which the model has seven, depicted as lorries.

3 STRAVVS versions 1+: Updating the theory

So far, we have a 'theory' and an implementation of it to prove its consistency, but we have not applied it to an empirical case study. There were two reasons for making changes, which interact with each other. The first reason is coding convenience and elegance; a matter that becomes more apparent once a serious application of the theory in an empirical case study, rather than the demonstrator in figure 2, requires more in the way of expressions accessing data in the model. An example of this is the reification of the *input* and *output* relations in figure 1 to *Input*, *Catalyst* and *Output* in the second version of the theory's ontology shown in figure 3. It is then more explicit that the *Processes*' operands are *Resource-*

types, rather than having distinct classes (*Make*, *Ingredient* and *Catalyst* in figure 1) for the roles the operands play in each *Process*.

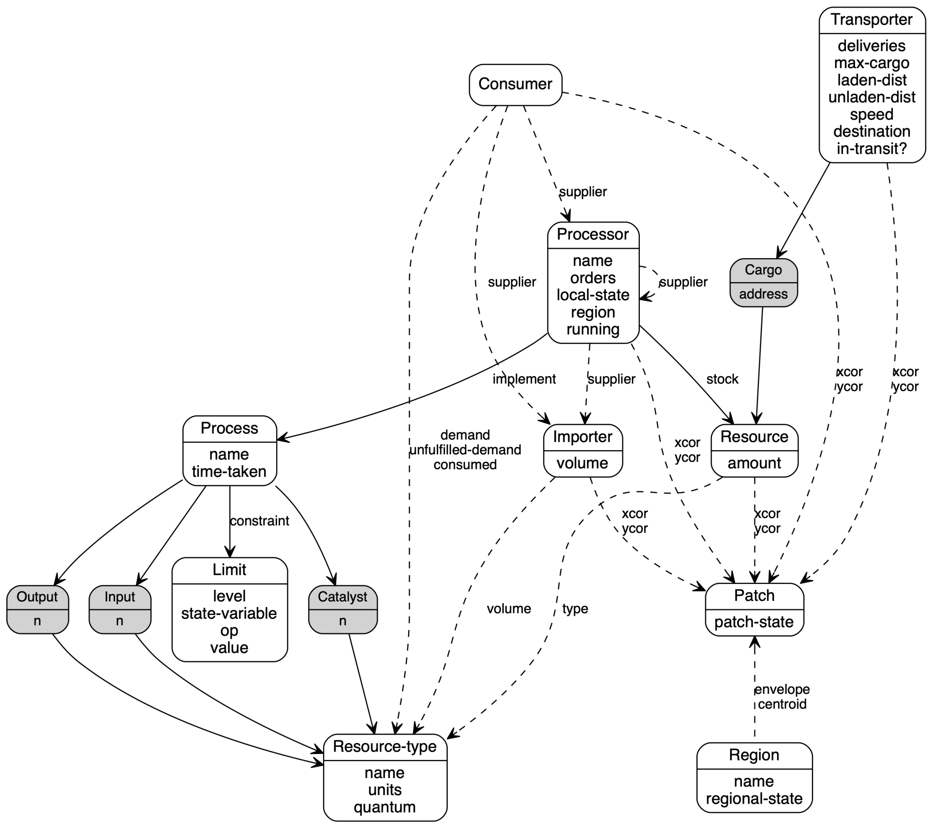

Fig. 3. Second version of STRAVVS ontology.

The second reason is to improve the ability of the theory to represent phenomena in the empirical world to enable the case study to be modelled with sufficient detail that it provides relevant and useful information to stakeholders in the model. A trivial example could simply be changing nomenclature to make matters clearer. Figure 3 uses *Resource* instead of *Product* in figure 1 because the latter wording implies that the objects in question have been ‘produced’, which is not necessarily the case. A more substantial example is the case where there are contextual *Limits* on where and when a *Process* might be run. This became most apparent as we developed the ‘equations’ (an input file to the STRAVVS model used to initialize *Processes* – see listing 1) to reflect the activities of (farm) businesses in the food security case study:

```
<equation-line> ::= <equation-name> " : " <equation>
<equation> ::= <catalysed-equation> | <non-catalysed-equation>
<catalysed-equation> ::= <inputs> " -[ " <catalysts> " ]> "
    <outputs> <time-limits>
<non-catalysed-equation> ::= <inputs> " -> " <outputs> <time-limits>
<inputs> ::= <resource> | <resource> " + " <inputs>
<catalysts> :: <resource> | <resource> " + " <catalysts>
<outputs> ::= <resource> | <resource> " + " <outputs>
<time-limits> ::- " | " <time> | " | " <time> " | " <limits-list>
<resource> ::= <resource-name> | <resource-multiple>
<resource-multiple> ::= <number> " " <resource-name>
<limits-list> :: <limit> | <limit> " & " <limits-list>
<time> ::= <number> " " <time-units>
<limit> ::= <string> " " <op> " " <value>
<value> ::= <single-value> | <value-list>
<value-list> ::= "{" <many-values> "}"
<many-values> ::= <single-value> | <single-value> " " <many-values>
<single-value> ::= <number> | <string>
<time-units> ::= <time-unit> | <time-unit> "s"
<time-unit> ::= "day" | "week" | "month" | "year"
<op> ::- "=" | "!=" | ">" | ">="  | "<" | "<=" | "in" | "!in"
```

Listing 1: BNF grammar for the 'equations'. Non-terminals `string` and `number` are not detailed for brevity, but neither may contain white space.

- *Processes* take *time-taken* to run. For example, the gestation time of cattle is approximately nine months.
- There are contextual *Limits* on *Processes*, pertaining to such things as environmental conditions needed to run the *Process* – crops have requirements for drainage, soil pH, temperature and rainfall during the growing season; and the machinery used will have requirements, especially for topological gradient. This means we need to add geography to the theory, with *Patches* and their agglomeration into *Regions*, and *-state* variables for the *Processor*, *Patch* and *Region* that can be checked by the *Limits*.

4 Discussion

The determination of an ontology for a model is a process often taken for granted, and yet more expressive ontologies is a central benefit of agent-based models [13]. As such, calibration and validation of agent-based models also entails consideration of their ontologies. Methods for calibrating and validating ontologies primarily involve assessment of 'interoperability', including the ability to instantiate or 'populate' the ontology in a specific case study. We have shown how considerations arising from applying the 'middle-range theory' we postulated in section 2 to an empirical case study in food security led to a need to modify the theory as shown in section 3. Motivations for updating the theory were derived from the need to simplify the implementation code and from unrepresented features of the case study important to modelling it successfully.

Further requirements to modify the theory have emerged since, perhaps the most significant being that the ontology as presented above contains no consideration of financial aspects. On the coding side, it is convenient for agents to be able to keep a record of what each of their *stock* is for, effectively meaning that relation in figure 3 is replaced by three: *input-stock*, *catalyst-stock* and *output-stock*. Those familiar with the farming industry will also be aware that the outcomes of activities have some uncertainty, which means the grammar for the equations in listing 1 needs updating so that the *Resources output* from a *Process* can be sampled from a distribution that may be parameterized not only by spatially-distributed features, but by events (e.g. unusual weather, diseases) during the *time-taken* of the *Process*.

In principle, we now have code that can take a set of equations as input, along with other supporting data to initialize the network of businesses, etc. and run a model simulating businesses exchanging goods, driven by demand for consumers (figure 4). This model *should* apply just as well to the provision of food as it would to the circular economy involving completely different materials (e.g. electronics), though we have yet to try configuring STRAVVS to a circular economy case study and discover further requirements for change. It is this that would prove a true test of the theory's generality.

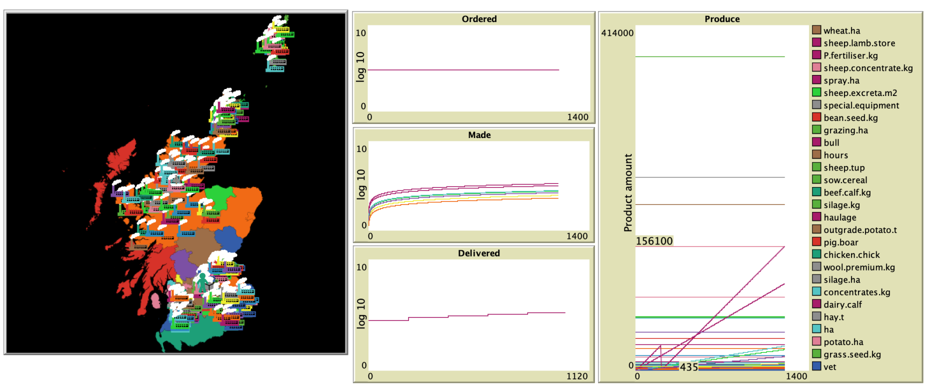

Fig. 4. Screenshot from a configuration of STRAVVS simulating food production in Scotland.

Evans et al. [5] make the case that simple theories in ecology risk being applicable to no real empirical ecological systems. The intellectual journey taken in the work described here supports that argument. The initial conceptualization would not be capable of convincingly modelling a scenario in support of evaluating potential policy interventions. Even as things stand, though we are carrying out a 'structural calibration' of the theory through applying it in one case study, we have not yet reached a point where we are validating the model's behaviour against the copious volumes of quantitative farm census data that would give us more confidence in the model's dynamics.

However, there are deeper questions to be asked. The ambition to undertake this modelling exercise as a theory-building task started with a seemingly elegant metaphor for the activities of businesses exchanging materials in a network that is driven by consumer demand. The intuition was that consumers are 'sucking' natural resources through the 'straws' of transformations in materials undertaken by successive businesses, and that this system-level overview of industrial economic activity would enable us to explore scenarios aiming to intervene in it to achieve policy ends. Applying this intuition to a specific case study has led to a number of complications in that initial conceptualization, and we have not yet reached the end of that process of complication. The central question, then, is whether it is reasonable to have embarked on a process of theorization in the first place, rather than writing separate empirical models to address each case, however much the likes of O'Sullivan and colleagues may YAAWN [12]? Can (a modified) STRAVVS become a middle-range empirical theory of the exchange of goods in a network of businesses that is useful for exploring the cascading consequences of policy interventions? Or are we, as has somewhat over-excitedly been argued for Big Data [1], finding our own arguments in social simulation for the 'end of theory' in the social sciences?

These questions aside, the intellectual status of scientific theories among researchers is such that the treatment of a model under development as a formal theory of the empirical world changes the mindset with which the task is approached. Changes to nomenclature, adding variables to classes (or so-called '`breeds`' in NetLogo), new relationships ('`link-breeds`') and new processes, though everyday tasks for a programmer trying to get code working, are seen differently when they are adjusting a theory. It is difficult to make general recommendations from a single example of a collaborative modelling exercise undertaken with the intent that the model be a 'theory', but reasonable to speculate that perhaps one of the reasons agent-based models do not lead to much theory development is that we do not treat them as theories. Provided we are willing to drop any preconceptions about properties of 'good' theories, such as elegance and simplicity [4], and make complications where needed to fit empirical applications, treating our models as theories of the empirical world could be one way to address a key criticism of agent-based models.

References

1. Anderson, C.: The end of theory: The data deluge makes the scientific method obsolete. Wired **23 June** (2008), https://www.wired.com/2008/06/pb-theory/
2. Balzer, W., Brendel, K.R., Hofmann, S.: Bad arguments in the comparison of game theory and simulation in social studies. Journal of Artificial Societies and Social Simulation **4**(2), 1 (2001), https://www.jasss.org/4/2/1.html

3. Boero, R., Squazzoni, F.: Does empirical embeddedness matter? methodological issues on agent-based models for analytical social science. Journal of Artificial Societies and Social Simulation **8**(4), 6 (2005), https://www.jasss.org/8/4/6.html
4. Edmonds, B.: Simplicity is *not* truth-indicative. In: Gershenson, C., Aerts, D., Edmonds, B. (eds.) Worldviews, Science and Us: Philosophy and Complexity. pp. 65–80. World Scientific, Singapore (2007)
5. Evans, M.R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C.M., Merz, M., O'Malley, M.A., Orzack, S.H., Weisberg, M., Wilkinson, D.J., Wolkenhauer, O., Benton, T.G.: Do simple models lead to generality in ecology? Trends in Ecology & Evolution **28**, 578–583 (2013). https://doi.org/10.1016/j.tree.2013.05.022
6. Gilbert, N., Ahrweiler, P., Barbrook-Johnson, P., Narasimhan, K.P., Wilkinson, H.: Computational modelling of public policy: reflections on practice. Journal of Artificial Societies and Social Simulation **21**(1), 14 (2018). https://doi.org/10.18564/jasss.3669
7. Jarvis, A.J., Jarvis, S.J., Hewitt, C.N.: Resource acquisition, distribution and end-use efficiencies and the growth of industrial society. Earth System Dynamics **6**(2), 689–702 (2015). https://doi.org/10.5194/esd-6-689-2015
8. Lorscheid, I., Berger, U., Grimm, V., Meyer, M.: From cases to general principles: A call for theory development through agent-based modeling. Ecological Modelling **393**, 153–156 (2019). https://doi.org/10.1016/j.ecolmodel.2018.10.006
9. Merton, R.K.: Social Theory and Social Structure. The Free Press, New York, NY, USA (1949)
10. Moss, S.: Game theory: limitations and an alternative. Journal of Artificial Societies and Social Simulation **4**(2), 2 (2001), https://www.jasss.org/4/2/2.html
11. Nagel, E.: The Structure of Science: Problems in the Logic of Scientific Explanation. Hackett Publishing Company Ltd., Indianapolis, Indiana, USA (1979)
12. O'Sullivan, D., Evans, T., Manson, S., Metcalf, S., Ligmann-Zielinska, A., Bone, C.: Strategic directions for agent-based modeling: avoiding the YAAWN syndrome. Journal of Land Use Science **11**(2), 177–187 (2016). https://doi.org/10.1080/1747423X.2015.1030463
13. Polhill, G., Salt, D.: The importance of ontological structure: Why validation by 'fit-to-data' is insufficient. In: Edmonds, B., Meyer, R. (eds.) Simulating Social Complexity (2nd edition). pp. 141–172. Understanding Complex Systems, Springer, Cham, Switzerland (2017). https://doi.org/10.1007/978-3-319-66948-9_8
14. Waldherr, A., Wijermanns, N.: Communicating social simulation models to sceptical minds. Journal of Artificial Societies and Social Simulation **16**(4), 13 (2013). https://doi.org/10.18564/jasss.2247