Social simulations: improving interdisciplinary understanding of scientific positioning and validity[1]

Stuart Rossiter (University of Strathclyde)
Jason Noble (University of Southampton)
Keith Bell (University of Strathclyde)

7th October 2011

[1]This is a post-print version of the article published 31 Jan 2010 with the following citation:

ROSSITER, S., NOBLE, J. & BELL, K. (2010). Social simulations: improving interdisciplinary understanding of scientific positioning and validity. *Journal of Artificial Societies & Social Simulation* **13**(1), 10. URL http://jasss.soc.surrey.ac.uk/13/1/10.html.

The definitive published version is available open access at the given URL in its default HTML format. This document represents a print-typeset version of the LaTeX original (before HTML conversion using the tools available on the JASSS site at http://jasss.soc.surrey.ac.uk/admin/submit.html). This may therefore be a more print-friendly version. (The JASSS site supports PDF downloads, but these are created directly from the HTML, and thus not typeset as this original.) *Because the published version uses a paragraph numbering scheme (with no page numbers), you should use that numbering scheme in any location specific citations.*

Abstract

Because of features that appear to be inherent in many social systems, modellers face complicated and subjective choices in positioning the scientific contribution of their research. This leads to a diversity of approaches and terminology, making interdisciplinary assessment of models highly problematic.

Such modellers ideally need some kind of accessible, interdisciplinary framework to better understand and assess these choices. Existing texts tend either to take a specialised metaphysical approach, or focus on more pragmatic aspects such as the simulation process or *descriptive* protocols for how to present such research. Without a sufficiently neutral treatment of *why* a particular set of methods and style of model might be chosen, these choices can become entwined with the ideological and terminological baggage of a particular discipline.

This paper attempts to provide such a framework. We begin with an epistemological model, which gives a standardised view on the types of validation available to the modeller, and their impact on scientific value. This is followed by a methodological framework, presented as a taxonomy of the key dimensions over which approaches are ultimately divided. Rather than working top-down from philosophical principles, we characterise the issues as a practitioner would see them. We believe that such a characterisation *can* be done 'well enough', where 'well enough' represents a common frame of reference for all modellers, which nevertheless respects the essence of the debate's subtleties and can be accepted as such by a majority of 'methodologists'.

We conclude by discussing the limitations of such an approach, and potential further work for such a framework to be absorbed into existing, descriptive protocols and general social simulation texts.

Keywords: social simulation, methodology, epistemology, ideology, validation

1 Introduction: Purpose & Goals

To set some overall context, figure 1 gives an abstract representation of the overall simulation process, emphasising the role of epistemological and methodological choices; compare the beginning practitioner view in Gilbert & Troitzsch (2005, §2).

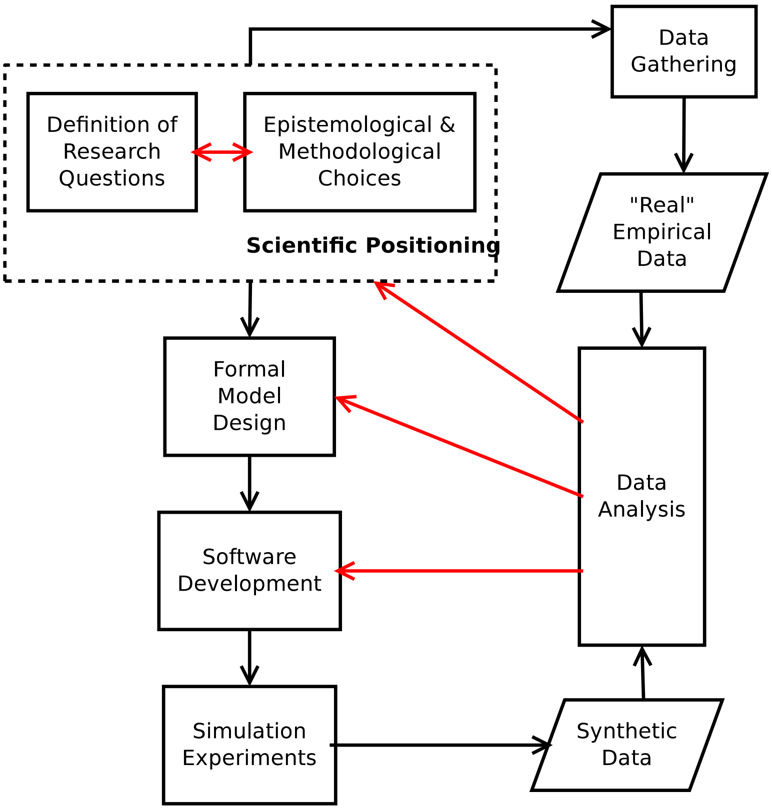

Figure 1: An abstraction of the overall simulation process. Black arrows show broadly sequential processes. Red arrows show the main feedback routes by which design decisions are adjusted

There is clearly iteration between the steps but, broadly speaking, decisions on scientific positioning come first. The researcher then progresses through formal (computational) model design, to software development and simulation experimentation (with empirical data gathering as required). Data analysis is the main mechanism by which previous design decisions are re-evaluated and the process iterated. (This is analysis both of real-world empirical data and the synthetic data produced by the simulation, often in comparison with each other.)

We will focus on scientific positioning, which includes:

- The nature of the research questions. This covers the *scope and purpose of what we want to know about the real world*.
- Epistemological and methodological choices. These cover *the ways in which modelling and simulation can give us scientifically valid knowledge about these research questions*.

It is crucial to note that there is *very strong interplay between these two elements*. Research questions suggest epistemological and methodological choices, but the latter also suggest particular ways of viewing the problem (and, indeed, the real world). This paper is concerned with *a framework which allows for an interdisciplinary characterisation of these epistemological and methodological choices*, the aim being to promote a 'universal' understanding of how social simulation research is scientifically positioned (the *'why?'* behind the research). Whilst the nature of the research questions is also part of this 'why?', understanding it typically presents little difficulty.

To justify our approach, we need to answer the following questions:

- What *are* the issues with positioning social simulation research in particular?
- How are they addressed in the current literature?
- What makes the approach here different, and how does it aid the field?

1.1 Scientific Positioning: Difficult Choices for Social Simulation

Epistemological and methodological debates in social system modelling are driven by the nature of social systems and the inherent difficulties in applying simulation as a formalised, computational approach.

The physical sciences are underpinned by universal mathematical laws, which typically allow for precise, quantitative matches to many real-world systems. The consistency of these laws also allows for strong predictive accuracy, and their relative simplicity means that Occam's razor is a useful measure in determining the validity of one theory over another; such **criteria of adequacy** are discussed by Schick & Vaughn (2007) and Chalmers (1999). Thus, the researcher in these fields is typically presented with fairly simple, objective choices of epistemology and methodology. We define such choices as one pole of an axis (figure 2), where moving towards the opposite pole reflects the increasing difficulty in making choices as we model more complex systems.

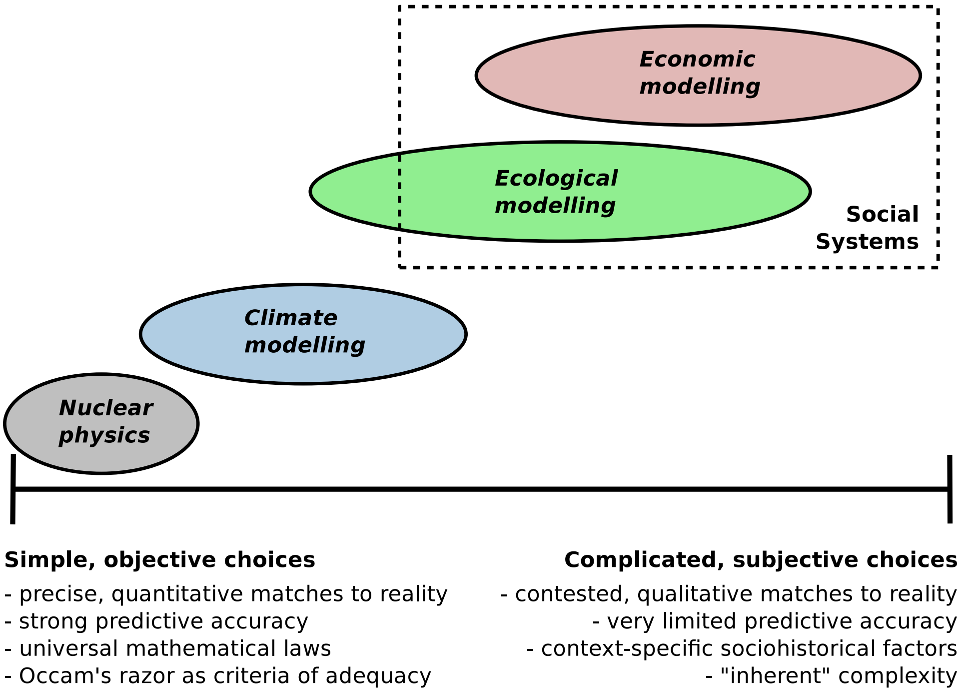

Figure 2: An 'axis of difficulty' for epistemological and methodological choices

Social systems can be regarded as potentially the furthest towards this second pole, for reasons as follows.

System Complexity. Nearly all social systems are profoundly complex, and some would argue *inherently so* for the purposes of modelling (Edmonds & Moss 2005). They are interconnected and not easily isolatable (e.g., markets for different products). Furthermore, they involve many interactions between different types of entitites (e.g., between individuals, or between individuals and structural institutions— the latter themselves a product of individual interactions). Because the participants are *human*, they are uniquely able to perceive structural aspects of the system and change or react to them, introducing complex *feedback effects* between humans and their environment[1].

Limited Accuracy. Primarily due to this complexity, and the fact that the system's structure and rules may be volatile over time, social system

[1]For a representative discussion, see Gilbert's article on structuration (Gilbert 1996). Of course, this is a very wide topic and also relates to debates on whether particular modelling techniques, such as agent-based modelling, may be better ways to reflect such complexity and how it influences macro-level properties of the system. This is reflected later in this paper.

models tend to lack broad accuracy; particularly predictive accuracy, but also quantitative descriptive accuracy. As Gilbert and Troitzsch put it:

> "[...] social scientists tend to be more concerned with understanding and explanation [than prediction]. This is due to scepticism about the possibility of making social predictions, based on both the inherent difficulty of doing so and also the possibility, peculiar to social and economic forecasting, that the forecast itself will affect the outcome." (Gilbert & Troitzsch 2005, p.6)

This helps perpetuate a range of ideological approaches since, to use Kuhn's view of science (Kuhn 1970), a dominant paradigm has not yet emerged (predictive accuracy being a primary means by which the superiority of a particular approach would be demonstrated). This has the follow-on effect that there is a proliferation of styles of model, with less efforts towards direct comparison and standardisation than in disciplines with more constrained methodologies; something which many social scientists would like to change (Axelrod 1997; Richiardi *et al.* 2006).

It also means that, in the absence of broad quantitative accuracy, empirical validation will be against qualitative patterns or selective quantitative measures. Debate therefore ensues over how best to select such data (Windrum *et al.* 2007; Moss 2008; Brenner & Werker 2007). There are also more fundamental questions over what type of mechanisms should be built into models (e.g., empirically-backed or not) to make such accuracy more likely; these questions can be intra-discipline (Edmonds & Moss 2005) or effectively define the boundaries *between* disciplines (e.g., the differences between economic sociology and neoclassical economics presented by Swedberg *et al.* (1987)). Much of this is caught up with social science's long tradition of advocating approaches which tend to reject the possibility of generalisable 'laws of human behaviour' and focus on the **sociohistorical** context of the social system (Eisenstadt & Curelaru (1976, §8.6) use the term "historical-systemic"). For modelling, this relates to the degree to which context-specific cultural factors affect behavioural patterns, as opposed to biological traits (Read 1990).

The lack of agreed axiomatic laws means that we potentially need a **model-centric** conception of social science, where *theory is represented by a family of models* (McKelvey 2002).

Issues with Empirical Data. Social systems are not easily isolatable, and the mechanisms by which they operate vary over time (so the social scientist may be limited by available historic data). In addition, many theoretical concepts are not easily formalised quantitatively such as, to use Bailey's example (1988), a person's sense of powerlessness. This engenders further debate into how experimental and data collection techniques are designed and validated (Bailey 1988).

1.1.1 Other Disciplines

As figure 2 suggests, such issues are not unique to the social sciences, and will generally occur in all areas where systems exhibit highly complex interactions (of individuals, structural entities, systems, etc.). This includes more directly comparable areas such as organisational theory and ecological modelling (particularly animal behaviour modelling), but also physical-science-based ones such as climate modelling. In the latter case, though the constituent elements (dynamics, radiation, surface processes and resolution) are strongly believed to be well identified and to operate according to fundamental laws of physics, the complexity of their interaction means that climate models retain a very short predictively accurate window, and only at restricted resolutions (Shackley *et al.* 1998).

Thus, we should not ignore the treatment of similar debates in these disciplines and, to that end, works from various disciplines are cited herein, without justifying each time why the ideas are relevant.

1.2 Our Approach in Context

We have shown the difficulties in making positional choices for social simulation. In particular, the ways to view and make these choices are typically entwined with the various disciplines and schools of thought which they engender; Windrum et al.'s discussion on validation techniques for agent-based models (2007) and Moss's response (2008) provide a good example. Given that such schools often have their own terminological baggage, it often becomes difficult to 'see the wood for the trees' and understand the common dimensions which differing approaches are opposed over (or perhaps agreeing on, but with different flavours of agreement). Fundamentally, *interdisciplinary assessments of scientific value and credibility prove difficult*.

This paper therefore attempts to provide a framework for a discipline-neutral understanding of these epistemological and methodological choices, which we believe is essential to improve the scientific assessment of social simulation research. Since social systems are just one flavour of **complex**

adaptive systems (CASs), and we draw on literature from other areas, our framework may have some wider applicability. (We discuss this further in section 4.)

Rather than working top-down from philosophical principles, we characterise the issues as a practitioner would see them. We believe that such a characterisation *can* be done 'well enough', where 'well enough' represents an *accessible* common frame of reference for all modellers, which nevertheless respects the essence of the debate's subtleties and can be accepted as such by a majority of 'methodologists'. (By referring to the more details debates, we also provide 'jumping-off points' for further study as desired.)

We begin with an epistemological model, which gives a standardised view on the types of validation available to the modeller and their impact on scientific value (section 2, based on a model-centric view of science). This is followed by a methodological framework, presented as a taxonomy of the key dimensions over which approaches are ultimately divided (section 3). We conclude by discussing some of the framework's limitations, and potential further work for such a framework to be absorbed into existing, descriptive protocols and general social simulation texts (section 4).

Elsewhere in the literature, such issues are usually discussed in one of four types of work, each of which we believe does not provide the accessible, interdisciplinary view that we are aiming for here:

Simulation textbooks. These, such as Gilbert & Troitzsch (2005), are generally aimed at the practical training of researchers new to simulation. Thus, they tend to focus on the general set of techniques (types of simulation, validation methods) and the software development process, with only an introductory or implicit coverage of more ideological issues[2].

Philosophy of (social) science. These tend to explore the broader metaphysical debates on how social science research can and should be conducted (Eisenstadt & Curelaru 1976; Burrell & Morgan 1979; McKelvey 2002; Gilbert 2004), such as contrasts between positivism vs. realism and etic vs. emic analyses (Gilbert 2004). This 'top-down' context is typically less useful for the average simulation researcher because: the discussion can be abstruse; and, by choosing to conduct simulation research, some epistemological choices have already been made (i.e., that a formal, computational model can provide useful knowledge on a real world system), which makes some of the debate superfluous and the remainder difficult to tease out.

[2]There is also the potential for the author, explicitly or not, to reflect their own ideological preferences— see Manzo's chapter 4 comments in his review of Gilbert's "Agent-Based Models" (Manzo 2008).

Specific methodological papers. These tend to focus on a particular issue— such as empirical validation (Windrum *et al.* 2007)— or a particular ideology's defence of its position in relation to another (Goldspink 2002; Brenner & Werker 2007; Moss 2008). Though these works give some useful abstractions, they do not give a sufficiently neutral global view, and often requires specialised knowledge of the particular philosophical points in question.

Pragmatic methodological protocols. These try to standardise how simulation model research is presented in the literature, and therefore have quite similar aims to our work here. Richiardi *et al.* (2006) compile a thoughtful and well-referenced protocol for agent-based social simulations. Since many of its points apply equally to any agent-based simulation, it is perhaps no surprise that this echoes similar attempts in other disciplines, such as Grimm et al.'s ODD protocol in ecological modelling (2006): this builds on previous heuristic considerations (Grimm *et al.* 1996; Grimm 1999), and was applied more recently to social models (Polhill *et al.* 2008).

However, the main point here is that these protocols focus on a *descriptive* classification of the discrete factors which need to be considered and disseminated, without consistently giving insight into *why* a particular combination of techniques might be chosen. In particular, without this 'why', there is a certain suggestion that the scientific value of a simulation study is correlated to the level of detail provided across *all* the defined areas— that is, more rigorous estimation, validation, and the like implies 'better' science. Yet, this is typically not the case. Certain approaches may largely ignore some points as irrelevant, whilst considering themselves no less scientific.

Without a complete context for the 'why?', the erstwhile modeller has only a limited conception of which points are more difficult to make a decision on, which are more entwined with decisions made elsewhere, and which are tacit 'dogmas' of their own particular discipline. Therefore, there is clear scope for such descriptive protocols to be merged with aspects of the framework presented here; we discuss this in the conclusions (section 4).

1.2.1 A Note on the Term Agent-Based

Richiardi et al. talk about agent-based models. Agency has no firm definition as such, but a much-cited one is that used by Wooldridge (2002), where the focus is on agents situated in an environment and autonomously able to react to it[3]. However, in the looser sense, agent-based models (ABMs) are often regarded as any model which explicitly models interacting individuals, typically with variation at the individual level. In ecology, the term individual-based models (IBMs) is used instead, emphasising that the focus is on genetic and phenotypic variation at the individual level, rather than other aspects of agency. (We reference several ideas from IBM modelling later.)

In this paper, we use the term **agent-based** in the wider sense, and interchangeably with **individual-based**, as Grimm also recommends (Grimm 2008). The choice of one or the other will generally depend on what literature we are discussing (and the terms used therein).

2 A Model-Centred Epistemology

If we are going to define a taxonomy to characterise the scientific positioning of social system simulations, we need to more formally define this positioning in the context of the scientific process. As McKelvey points out (McKelvey 2002), demonstrating how the wide range of approaches can fit within a single scientific epistemology is a non-trivial task, there being approaches which appear to reject aspects of the traditional 'scientific method' in varying degrees.

We build on an epistemology presented by McKelvey (2002) and Azevedo (2002) from organisational science (there are slight differences between their approaches, but the common core is what we are interested in)[4]. It is based on the semantic conception developed by philosophers such as Suppes, Suppe and Giere, but the point is that this can be shown to serve our purposes in practice, whether or not some approaches in the field are inspired by slightly different philosophical principles.

This epistemological model defines concepts and terms related to the various forms of **validation** within the modelling process, which are important in understanding the methodological taxonomy later (section 3).

2.1 Core Framework

The key features of McKelvey and Azevedo's framework are as below, and are summarised in figure 3 (which compares it to the 'Newtonian' axiomatic view).

[3] The meaning of 'autonomous' for a deterministic software system remains ambiguous, as noted by McArthur *et al.* (2007).

[4] Goldspink (2002) also builds on McKelvey's model but, unlike us, does this to advocate a specific (situated research based) methodology.

Science as Explaining the Dynamics of Phase Spaces. The quality of a science is governed by how well its models explain the dynamics of the phase space of the system in question (i.e., the space of all dimensions of the system). This could be at the level of predicting qualitative changes, rather than quantitative accuracy of detail (the latter tending to be the aim of the axiom-driven physical sciences).

Theories explain Isolated, Idealised Systems. There is never complete representation and explanation of the system, but of an abstracted one which, nevertheless, provides "an accurate characterization of what the phenomenon *would have been* had it been an isolated system" (McKelvey 2002, p.762, quoting Suppe).

Model Centredness. A *set* of models represents the theory, and this set will typically explore different aspects of the system in question (using different abstractions). There is *not* necessarily any definitive axiomatic base, though this is not precluded for some or all of the set of models. As McKelvey puts it: "Thus, 'truth' is not defined in terms of reduction to a single axiom-based model".

Put simply, *models are first class citizens of science*.

We view this is an intuitively 'correct' representation of complex systems science, that helps orient modellers regarding the 'point' of social simulations, even if their model focuses only on a particular aspect of the real-world system.

Azevedo (2002) provides a simple, yet powerful, analogy for this model-centred epistemology, which can be a useful summary for the detail above. She sees models as *maps*. Different types of maps are appropriate in different situations (e.g., a contour map for terrain analysis, or a symbolic map of key checkpoints for journey planning); equivalently, models may serve different purposes (e.g., state transition prediction versus detailed quantitative prediction within a state). A suitably specialised map can be *more* useful than the actual, physical area itself for a particular problem; equivalently, a particular model abstraction might cleanly represent one particular real-world aspect better than a massively detailed 'reconstruction' of the real thing.

2.1.1 Model Usage and Adequacy Testing

More specifically, McKelvey introduces some very useful concepts on the types of validation the researcher might aim for and, citing Read (1990), how the use to which they put their model reflects this (and can be characterised).

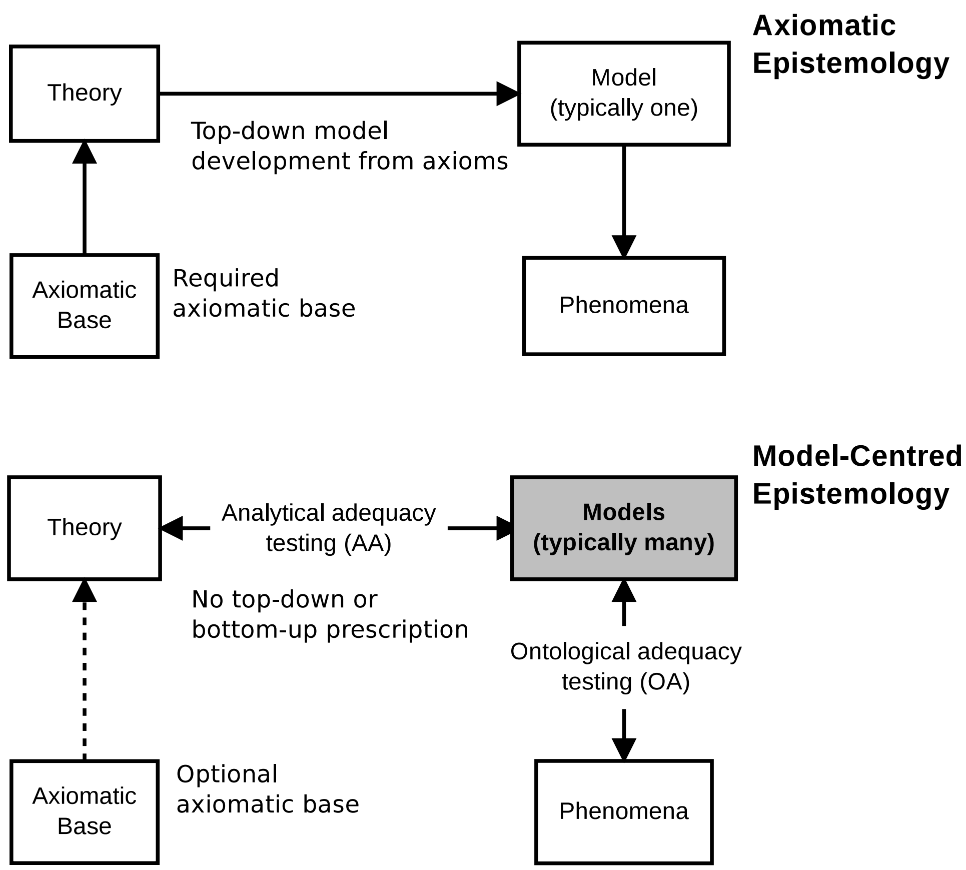

Figure 3: Comparing an axiomatic epistemology with a model-centred (semantic conception based) one— McKelvey and Azevedo's core position (diagram as per McKelvey (2002), with extra explanatory text and model shading to indicate its key position)

Models can be used as concrete representations of theory (in Read's terminology, a Model$_T$ usage). When used in this way, the research focuses on the exploration of (complex) theory and its potential consequences for the isolated, idealised system in question (without having to necessarily predict real-world behaviour); it can potentially show *dissonance* within the current theory and suggest potential changes of detail or direction. This is effectively a form of **theory–model validation**, and McKelvey refers to it as **analytical adequacy testing**.

A good example would be Schelling's famous segregation model (Schelling 1971). He uses a simple cellular automata to show that strong racial segregation can occur with only *mild* individual racial preferences. Thus, the 'point' was to question whether existing theory was really looking at the issue in the right way, and to alert theorists into the possibilities of emergent, system-level behaviour which is unintuitive given the individual-level rules.

Models can also be used to represent processes which reproduce (*describe*) some aspect of real-world empirical data (Read's Model$_D$ usage). Schelling's original research might have *suggested* such a use, but its primary purpose was theoretical. If some other piece of research determined empirical values for individual preferences, and then used the model to predict the system-level pattern in some way, then *this* research would be a Model$_D$ usage. Thus, such a usage will tend to focus on statistical techniques to determine goodness of fit. This is effectively **model–phenomena validation**, and is referred to by McKelvey as **ontological adequacy testing**.

Equally, a statistical regression fit to data represents a model with this Model$_D$ usage, but one which has been developed bottom-up from data, rather than top-down via posited theoretical mechanisms. As Read notes (Read 1990, p.34), such research still has *some* theoretical basis (for why this particular statistical model was deemed applicable), but the research is interested only in whether it fits the data, not exploring its theoretical origins or consequences (if this *was* explored, this part of the research would be Model$_T$ usage).

This demonstrates that the usage is a property of the *piece of research*, not the model per se (hence the emphasis on 'usage'). We will use the terms **theoretical model usage** for Model$_T$ usage, and **descriptive model usage** for Model$_D$ usage. To avoid awkward prose, we also sometimes say that a model *'is'* a theoretical model, meaning that it is being *used* as a theoretical model in the particular context we are discussing. Although the terms 'theoretical' and 'descriptive' are very general, we have found them to be the ones which most closely reflect the exact distinction, which is to do with the epistemological purpose of the modelling research, not how it is derived[5].

[5]For example, bottom-up vs. top-down refers to the *method*, not the *purpose*. It also has the dual meanings of 'from data' vs. 'from theory' *and* 'generating emergent behaviour' (as per Epstein & Axtell (1996)) vs. 'system-level rules'.

In ecology, Roughgarden *et al.* (1996) use the concepts of 'minimal models for ideas', 'minimal models for a system', and 'synthetic model for a system', which represent increasingly descriptive uses of the model. This is much more like what we want, but tends to refers to the model itself, not the research using it. There is also the idea of *progression* here which we are not comfortable with. For example, one could have a model which replicates an existing model that closely compares to real-world data (i.e., this other model is used in a strongly descriptive way). However, the purpose might be to show that the original modellers had overlooked a sensitivity to some crucial parameter which affected the plausability of their empirical matches. Thus, the new research is really a theoretical use of the model, not a descriptive one, since it doesn't offer alternatives to fix the empirical accuracy problems.

Notice that analytical and ontological adequacy tests make the very useful division between theory–model and model–phenomena validation. It is worth emphasising that typical empirical tests (e.g., a statistical fit against some particular measure) are *not* just ontological adequacy tests: they are effectively testing a combination of analytical and ontological adequacy, since the researcher cannot separate whether any lack of fit is due to the model being a generally inappropriate one (ontological adequacy issues), or that some invalid formalisation was made in transitioning from theory to model (analytical adequacy issues). Stanislaw (1986) discusses this, and the resultant need for specialised tests which *can* isolate a particular type of adequacy test: for example, experimenting with small structural changes to the model to determine whether certain decisions in formalising the model may have significant effects on the behaviour (analytical adequacy).

Therefore, both types of adequacy may be explored in parallel. A model which can be shown to fulfill both theoretical and descriptive uses will be a useful piece of empirical science but, crucially, there is still scientific value in each usage type taken separately: *theory–model and model–phenomena research are separate and equally viable scientific endeavours*.

2.2 Extensions for Social System Simulations

There are some useful extensions that we can make to capture the particular epistemological context for *social system simulations*. Each of these is explained in the sections which follow, and each can be represented diagrammatically in a simple way, building cumulatively from Figure 4, which represents McKelvey's model-centred base in stripped-down form.

The additional features added each time are marked in red on the revised diagrams. Note that we are only looking at the epistemological concepts here; the positional taxonomy will expand further on how particular methodological approaches in the literature map to this framework, and what aspects they choose to focus on.

These extensions define some new validation types, which also allows this epistemological model to be more directly compared to other validation frameworks (Stanislaw 1986; Bailey 1988). We believe that our model offers certain advantages, but the argument is more of interest to methodologists and so is left to appendix B.

2.2.1 The Importance of Model Usages and Usage Transitions

The modeller faces an important choice on which mix of theoretical and descriptive model usage their modelling research is going to explore, which

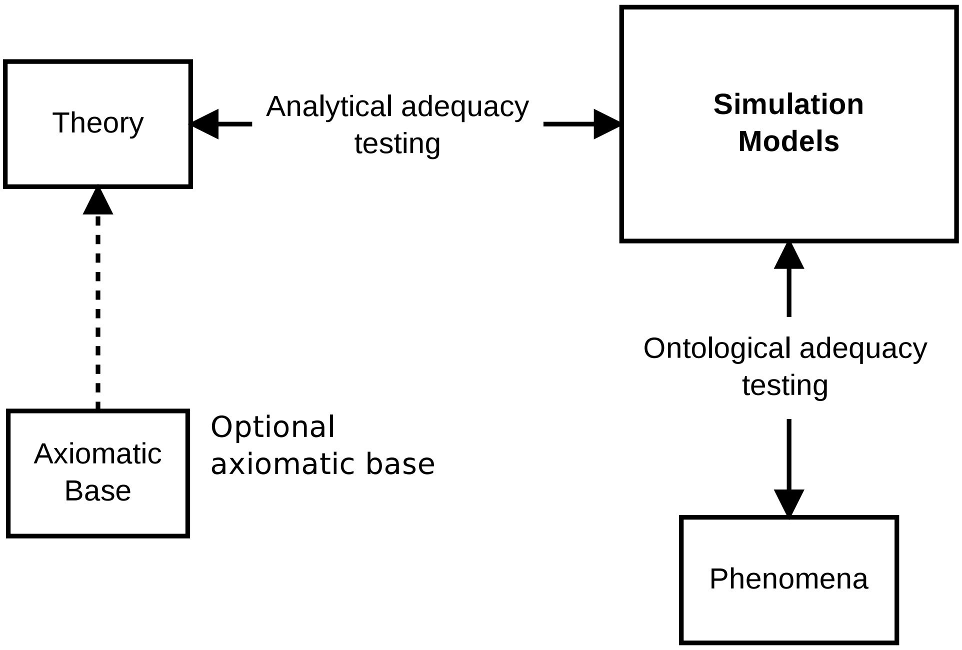

Figure 4: The base McKelvey model, ready for the addition of our social simulation amendments

we explore further in the methodological taxonomy (section 3); in many ways, this choice is an *iterative* one, influenced by the results from simulation experiments.

In addition, the arguments by which modellers move *between* the two usages is of key epistemological interest. Figure 5 shows this.

The transition from theoretical to descriptive usage is what we will call the **bridging argument**, as used by Voorrips (1987). This argument is the formation of a hypothesis on how some aspect of the real world works (e.g., by analogy from theory in another field). Its validity in itself will be related to subjective assessment by criteria of adequacy such as testability, simplicity and conservatism (Schick & Vaughn 2007). Such an argument is often made implictly; for example where a particular theory is intuitively postulated from examination of the empirical data, without any more formalised consideration of alternatives or the theoretical context. To paraphrase Read's example from Hill's archaeological research (Read 1990) , statistical relationships were explored between the spatial position of some remains and their type (e.g., storage jar). The implicit theory was that particular rooms were used for particular purposes and so, assuming little depositional disturbance over time, there would be a correlation. However, the focus was on corroborating the statistical fit and not on jus-

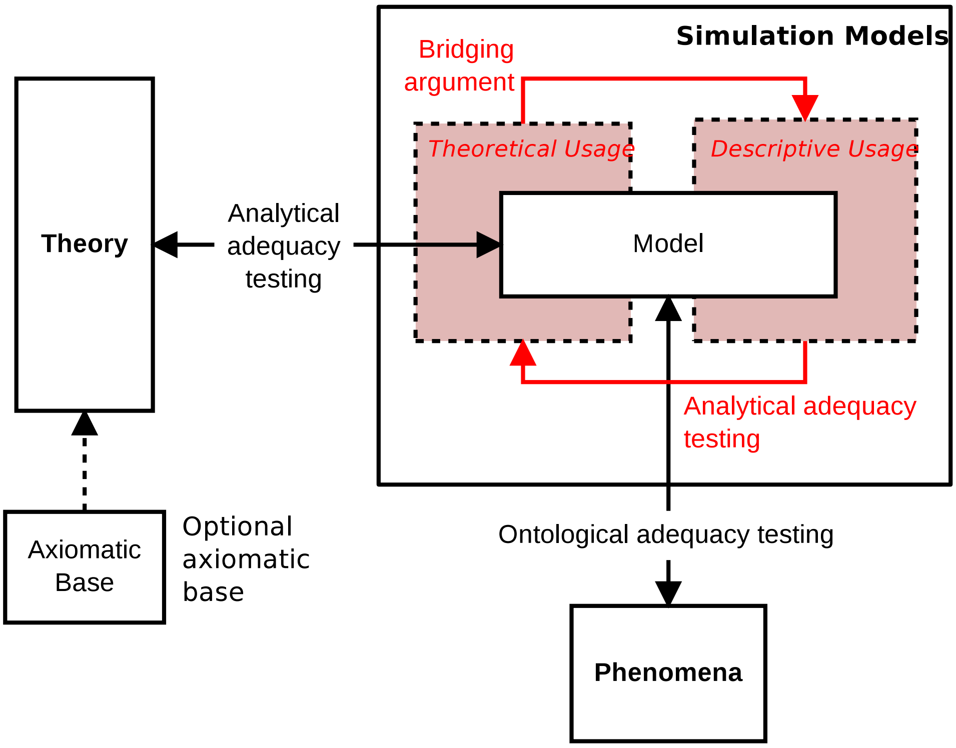

Figure 5: Adding concepts for model usages and usage transitions

tifying the theoretical premises or, for example, considering rival theories which might also result in such a correlation.

The reverse transition (descriptive to theoretical usage) involves the positing of theoretical mechanisms which can be shown to result in the regularities observed in the original descriptive model, and which fit into the existing theoretical context. This is therefore part of the theory–model validation (i.e., analytical adequacy testing).

2.2.2 Better Reflect Status of Empirical Data

McKelvey's view shows models being validated against phenomena. We noted in our introduction that there are extensive debates regarding the use of empirical data in model formulation, and empirical data's relationship to the underlying real-world system. In particular, the choice of empirical data is never truly objective, in that it is influenced by theoretical considerations and biases[6].

[6]This is a fundamental issue in the philosophy of science, as discussed by Chalmers (1999). Despite attempts by the 'new experimentalists' (e.g., Mayo) to ground science on an objective set of severely tested experiments, the experimentally-driven ap-

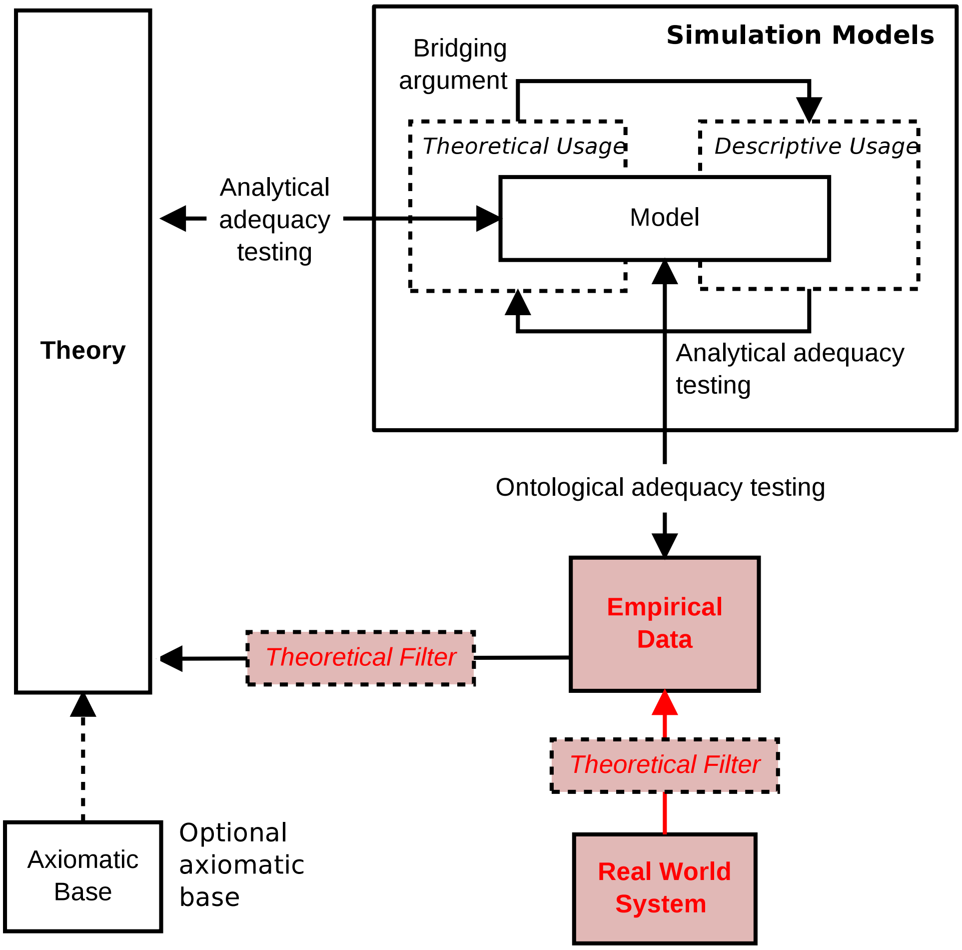

Figure 6: Adding concepts for the status of empirical data

This motivates some further refinements, shown in Figure 6:

- We reflect the distinction between the real-world system as a data-generating process and the observed empirical data that it generates (Windrum *et al.* 2007).
- Empirical data is linked to theory, to reflect that it is used in the formulation of theory and hence (indirectly) the model. Since the choice of what empirical data is appropriate, and the role of a priori assumptions, is ideology dependent, we show this as a *theoretical filter* on this use of empirical data.
- Similarly, a theoretical filter is applied to the observation process by

proach remains influenced by theoretical considerations.

which empirical data is obtained from the real-world system.

2.2.3 Computational Models and Software Adequacy

Figure 1 illustrated that the modeller has to move from a formal model to an actual software implementation. As well as the software engineering, this may also involve developing or using algorithms to approximate the required mathematics of the formal model (e.g., computing an integral by numerical methods). This process also has to take account of potential numerical issues which may produce artifactual system dynamics, perhaps due to precision errors— see, for example, Polhill et al., as discussed by Edmonds & Moss (2005, §3).

For our purposes, the point is that this is really a separate type of validation (see figure 7). We define this new validation as **software adequacy testing**, since it is all directly related to validating the adequacy of the software representation, with respect to the conceptual (mathematical) model. Hence, we can also refer to it as **conceptual–computational model validation**. To accomodate this definition, we more strictly define analytical adequacy testing as related to testing the relationship between theory and a *formal conceptual model*.

2.2.4 Explanatory Ability and Causal Adequacy

Descriptive accuracy does not necessarily imply that a model is a valid *explanation* of phenomena. Essentially, a causal explanation has to have other evidential support for its mechanisms being the 'real' ones, since just replicating empirical data is a weak argument: infinitely many other systems could potentially generate the same data. Thus, ontological adequacy does *not* cover this required validation. We also cannot take it as included in analytical adequacy, since this focuses only on the model as a representation of the theory.

Therefore, we need a new type of adequacy test. We define **causal adequacy testing** as between theory and empirical data from the real-world system (see figure 8). Note that this occurs *outside the simulation process*— the simulation only tests the *descriptive* accuracy of the model, not whether the model's mechanisms can be shown to actually work like that (individually) in the real world.

Grüne-Yanoff (2009) emphasises why this evidence is difficult for social systems: there are often 'fuzzy' social concepts which are difficult to formalise and observe in the real-world system (to empirically confirm causal adequacy). He points out that there is a weaker position that the simulation is a *potential* causal explanation. However, this is also problem-

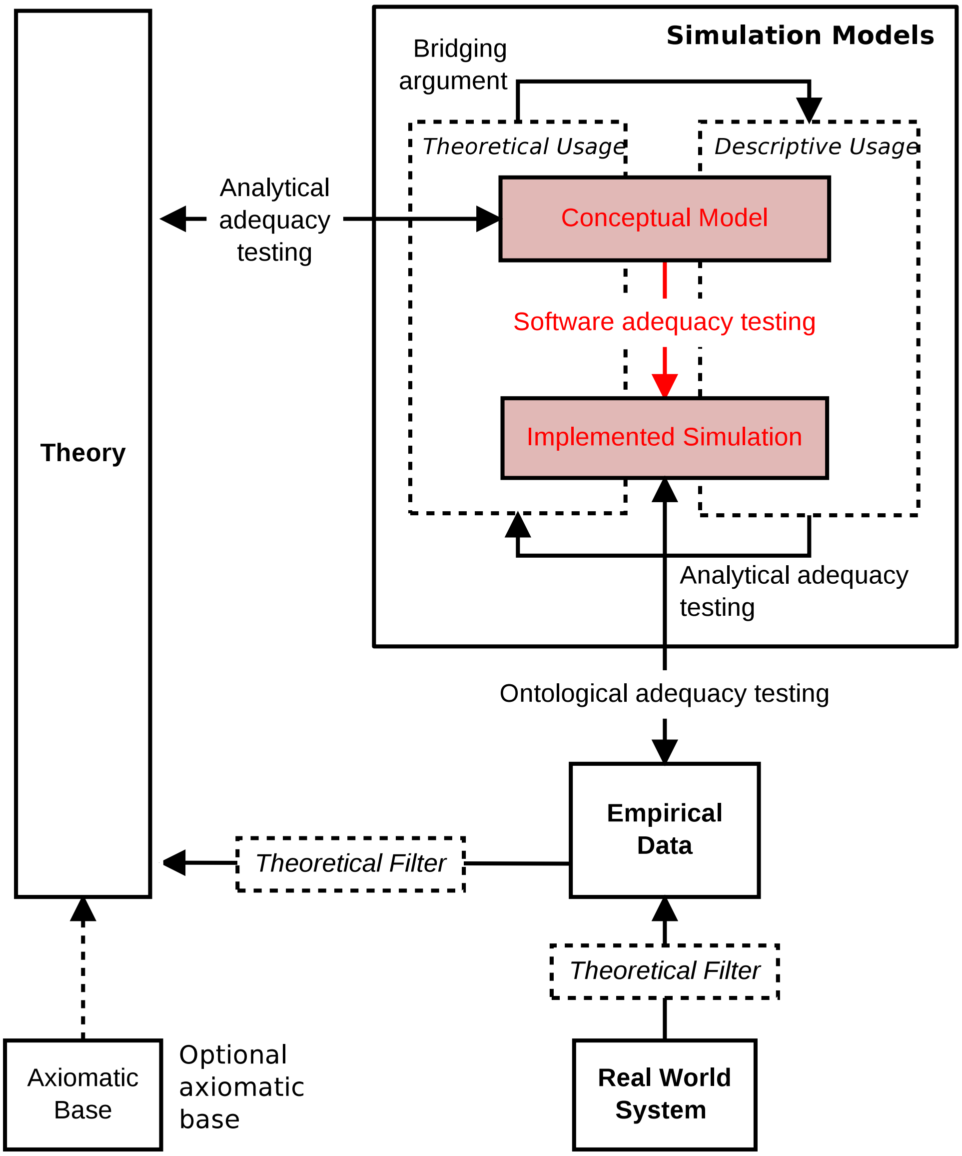

Figure 7: Accounting for the computational nature of the model with software adequacy

atic for social models because they often have many degrees of freedom, and can thus potentially reproduce empirical data with many parameter variations (so they do not narrow down the potential explanations very far); a lack of wider predictive accuracy is also an indicator that any accuracy for particular case studies may be such 'data-fitting', rather than true explanation.

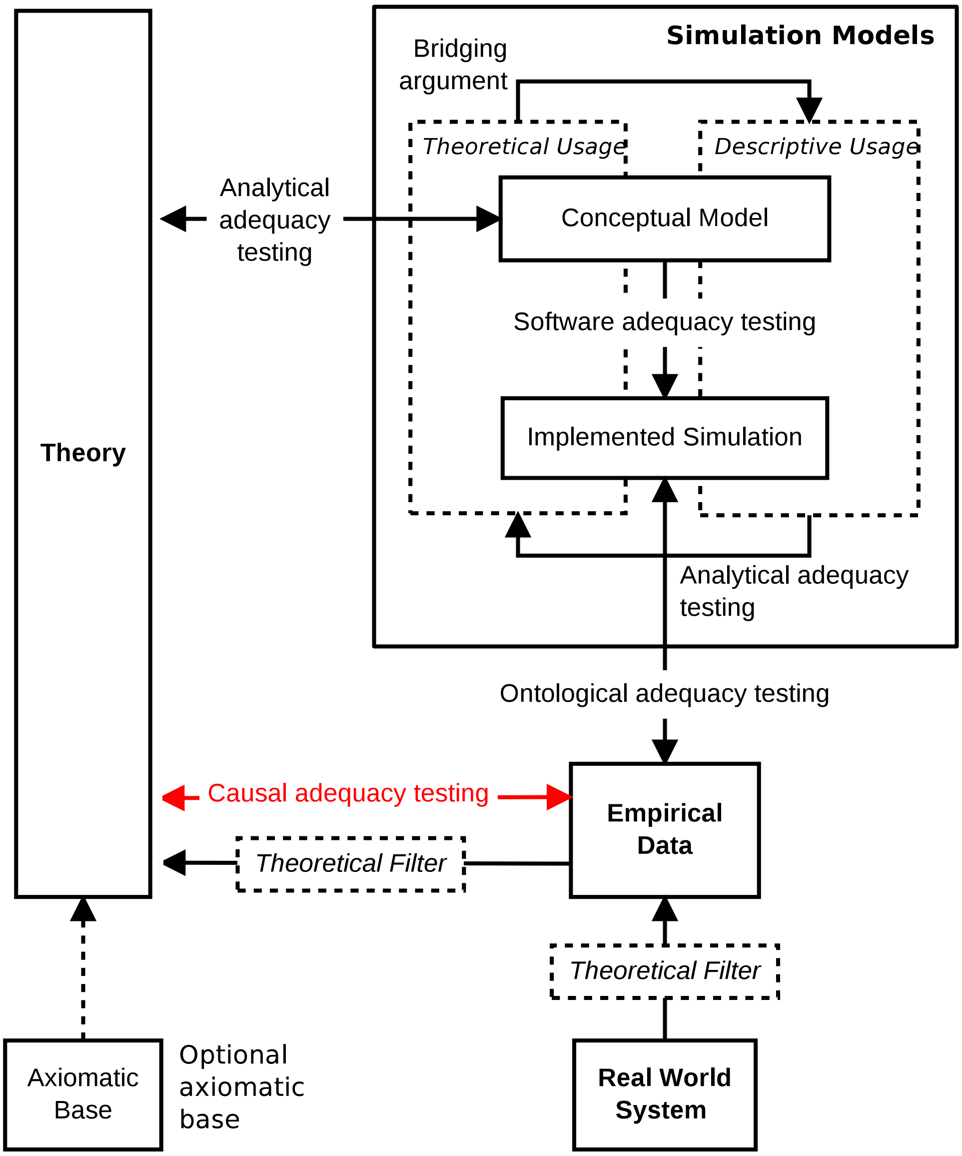

Figure 8: Adding the need to validate the proposed causal mechanisms against reality— causal adequacy

2.2.5 Reflecting Stakeholder-Centric Approaches

As a particular variant of the sociohistorically-oriented approaches discussed earlier, there is a significant body of social simulation research which uses system participants and external experts as central to the model design and testing process: what we will call a **stakeholder-centric** approach. The principal aim is to reach a shared understanding of, and belief in, the workings of the model and what it shows.

In a policy-making context, this is related to the pragmatic view that the model can only be useful if the stakeholders believe it and feel that it represents their views accurately. Such an approach is traditionally taken by the Operational Research (OR) community, for problems where the concern is with social/business interactions rather than manufacturing or logistical issues; the latter are problems where there often *are* standard theoretical models which predict the real-world system well (e.g., queueing theory for assembly lines). Howick et al.'s studies of change management in large projects (2008) fall squarely into this category, with efforts to better integrate more participatory approaches into the modelling process via their modelling cascade methodology.

In social systems modelling, the companion modelling approach espoused by Barreteau *et al.* (2003) more directly enshrines this stakeholder centricity as a fundamental tenet:

> "Instead of proposing a simplification of stakeholders knowledge, the model is seeking a mutual recognition of everyone representation of the problematique under study." (Barreteau *et al.* 2003, §4.3— sic)

This is a particular flavour of the view that theory is represented by a *set* of models which explore different abstract, idealised systems: in this case, the abstraction is stakeholder-centric. As Moss points out, this means that such models may only be valid in the context of the stakeholders that they are informing. Whilst all sociohistorical approaches imply constraints on how general a class of systems their theory is likely to apply to, stakeholder-centric ones also constrain our system definition to the system *as perceived and agreed subjectively by the stakeholders*; the principal aim becomes the addition of formal precision to debate, not the accurate forecasting of future behaviour (Moss 2008).

This motivates a revision to our homogeneous definition of theory by adding an **objective–subjective** dimension in figure 9. This is a useful pragmatic distinction in understanding what knowledge the model is trying to capture and potentially explain. (There are potentially some deeper philosophical issues with the concept of 'objective' and 'subjective' theory, but this is beyond the scope of this paper, and does not outweigh the practical usefulness of the terms.) Note that we are not using 'generalised–contextual' or similar, since this can be confused with the more general constraint on the theory's scope of application[7].

[7] This is often closely linked to the 'subjectivity' of the theory, but is not the same thing.

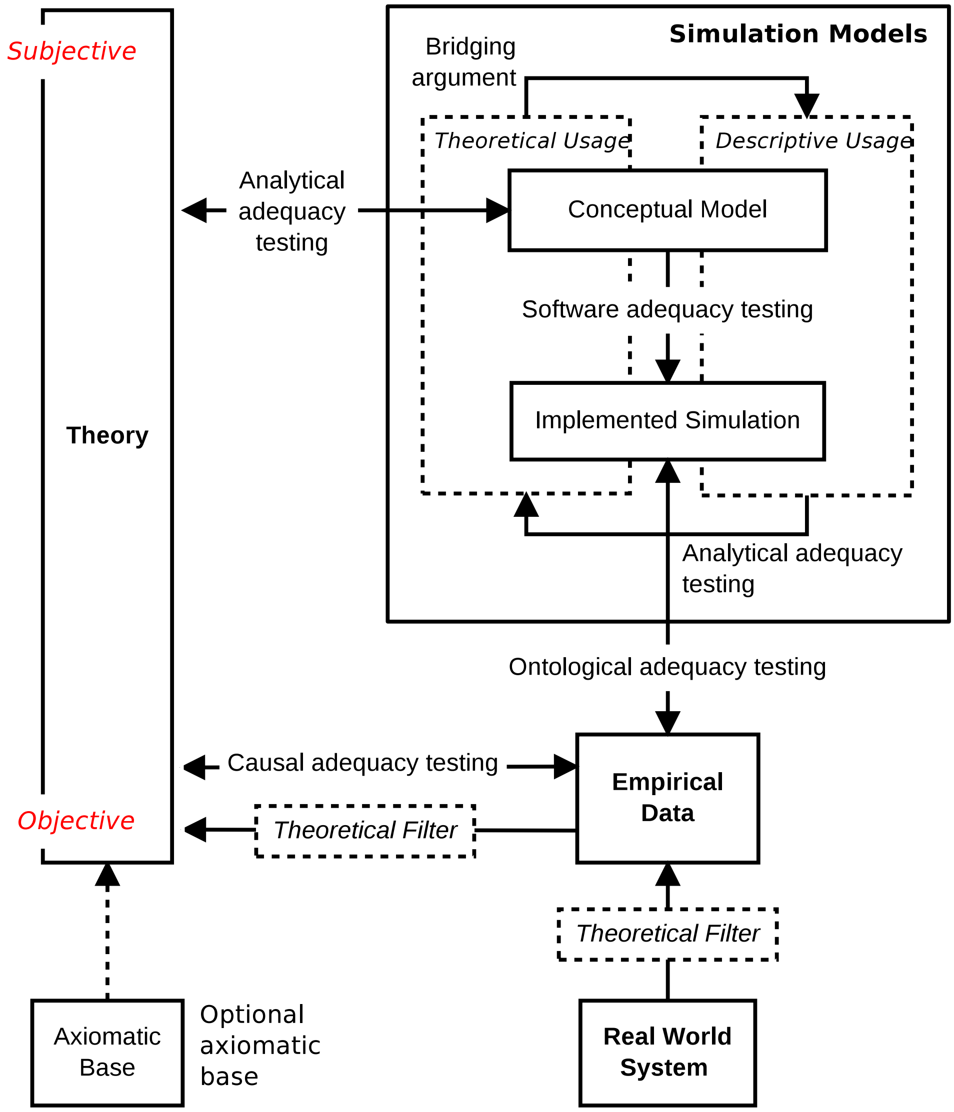

Figure 9: Incorporating stakeholder-centric approaches: considering theory as objective or subjective

3 Positional Taxonomy

The aim here is to present a taxonomy which reflects the essential, high-level issues which underpin epistemological and methodological decisions

A stakeholder-based model is often limited in its applicability to the particular system that the particular stakeholders are part of. However, there may be other stakeholders in the same system who were not included in the modelling, and so it will not necessarily represent *their* views.

when creating social simulation models. The categorisation is via a set of dimensions, each of which is a continuum of positions between two extremes (poles). Attempting to tease out these categories from debates in the literature involves both:

- the aggregation of largely independent debates which we argue are really different aspects of, or responses to, more fundamental ideological questions;
- the identification of the core methodological principles underlying general approaches— e.g., KIDS[8] (Edmonds & Moss 2005) or abductive simulation (Werker & Brenner 2004)— and the confirmation that these can be suitably characterised by the taxonomy.

One useful heuristic is that, when the classification is applied to a large number of approaches in the literature, the positions on the various axes should be largely independent; i.e., if a position on one dimension correlates with a position on another most of the time, that probably means that those dimensions do not represent distinct enough ideological issues.

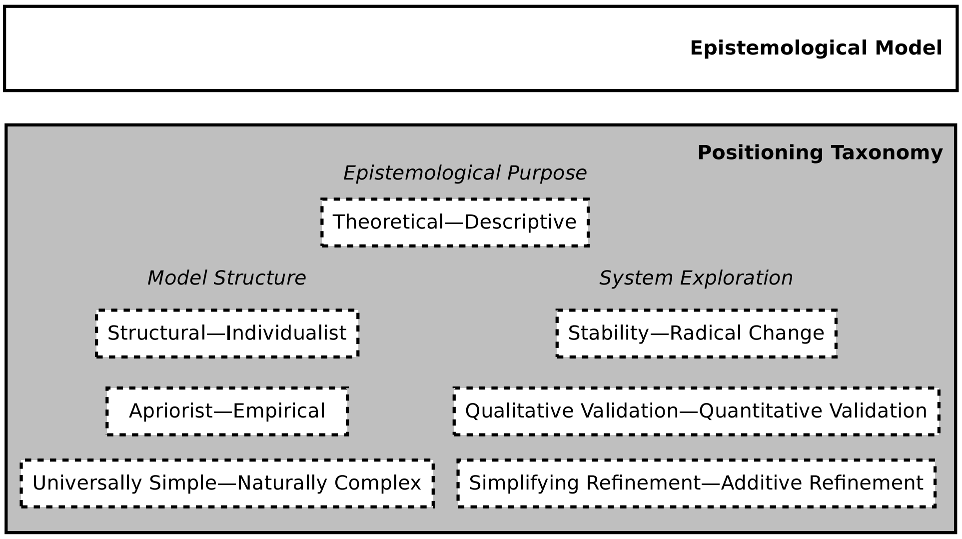

Figure 10: An overview of the proposed taxonomy

The taxonomy is summarised in figure 10. We have already shown that the decision regarding the mix of theoretical and descriptive model uses is particularly key for establishing the epistemological purpose of the

[8]Keep It Descriptive Stupid. Note that this is *not* descriptive in the sense of our descriptive model usage; Edmonds and Moss's use of the term is related to the *Apriorist – Empirical* and *Universally Simple – Naturally Complex* dimensions (see later).

research. Hence, this forms the most fundamental dimension (*Theoretical–Descriptive*), which will tend to govern and be influenced by decisions on other dimensions. These other dimensions split into two natural domains: how the model is structured, and how it is explored through experiment (simulation runs).

Each dimension is explained in the sections which follow. To help confirm the applicability of the taxonomy to the range of approaches mentioned throughout this document, appendix A provides a concise summary of each approach's position on the various axes.

3.1 Theoretical – Descriptive

The basic concepts of theoretical and descriptive usages are covered at length in the epistemological model (particularly with relation to how a model can be both theoretical and descriptive, and the bridging argument between such uses). This is something of an 'odd one out' dimension, because it is not really a dimension: a piece of research could have strong theoretical *and* descriptive model uses, though this would be unusual.

However, there are some practical subtleties which a modeller needs to be aware of when making this choice.

3.1.1 The Iterative Nature of the Choice

This choice is the one which can most change during the course of the research, because it is strongly governed by the *outcomes* of simulation runs, and how they cause the researcher to re-assess the nature and purpose of their model (refer back to figure 1).

A piece of simulation research which starts out as an attempt at empirical accuracy may end up being more useful as a theoretical result. A classic example would be Lorenz' attempt to model simplified weather systems via convection equations (Lorenz 1963). This ended up being a theoretical result (and presented as such), since it transpired that his model showed unusual effects such as sensitive dependence on initial conditions and bounded, yet non periodic, solutions. That is, it showed hitherto unseen dynamical properties of a simple, nonlinear system which also served to show the potential problems in trying to use the model descriptively for predictive accuracy. The paper turned out to be important in helping define chaos theory, and it was only later that researchers realised that such issues could also be *useful* for descriptive model uses, since they could be used to explain apparently complex behaviour via simple equations.

Because of this, it is important for the position on this axis to be clearly stated. Simulation tends to have an ability to seduce with its re-enactment of a model in dynamic operation, especially for agent-based models where the defined individuals and their rules tend to have an 'immediate believability' (unlike, say, a set of differential equations). Bullock highlights such issues for artificial life (ALife) modelling, which tend to be theoretical models that can sometimes be erroneously interpreted as suggesting direct descriptive uses:

> "In addition, there is little explicit work on combatting the downside of a simulation model's immediacy — the tendency of some audiences to 'project' added reality onto a simple simulation, mistakenly understanding the superficial similarity between simulated agents and real organisms as the point of a model, for instance." (Wheeler *et al.* 2002, §3.1)

3.1.2 Stylised Facts as Descriptive Models

We explained previously that, in the absence of broad empirical accuracy, much social systems research has to focus on *qualitative* accuracy in reproducing important patterns in the data: these patterns are often called 'stylised facts' in the literature. This is discussed in more detail within the *Qualitative Validation – Quantitative Validation* dimension, but there are a couple of points relevant here. We use Sallans et al.'s agent-based model of integrated consumer and financial markets as an example (2003). Their paper focuses on showing that their model can reproduce some key stylised facts.

Firstly, we should note that stylised facts *are* descriptive models by definition (though not simulation models): they formalise patterns in the empirical data without providing any explanatory mechanisms for them.

Secondly, each of the stylised facts may have been derived by a bottom-up approach (statistically inferring patterns from data) or from a *separate* theoretically-derived model which was also shown to act as a valid descriptive model. Sallans *et al.* (2003) has examples of both:

- For the former, high price volatility ("price volatility is highly autocorrelated. Empirically, market volatility is known to come in 'clusters'."): this is something observed empirically by market analysts. (It may also be predicted by other theoretical models, but we assume that it was first observed 'on the floor'.)

- For the latter, low predictability of price movements: it is an outcome of the efficient market hypothesis (a theoretical model) which was, to some degree, borne out by empirical data.

Thirdly, research which attempts to match to stylised facts (such as Sallans et al.) is *not* necessarily just a descriptive use of the model. As they state, their longer-term aim is to use such a model to explore the mutual interaction between the two markets, where this paper lays the groundwork:

> "Before we can use the model to investigate inter-market effects, we have to satisfy ourselves that it behaves in a reasonable way." (Sallans *et al.* 2003, §1.5)

This is therefore a mix of theoretical and descriptive uses, with the overall aim being more of a theoretical one: it attempts to show that their agent-based theory provides an alternative theoretical approach to conventional microeconomic models (one that produces types of behaviour which cannot easily be reproduced by conventional models). If agent-based modelling was a more established paradigm, and the particular agent definitions used had been researched separately, then the theoretical usage of the research would likely be more negligible. As it is, there is an inherent need to justify the approach and contextualise it within prevailing theory. It is used descriptively to match the stylised facts, but this is in order to demonstrate its plausability for further research on other aspects of the market.

3.1.3 Distinct Types of Theoretical Model Usage

Theoretical model uses allow us to explore the theoretical consequences of some system model. However, it is important to be clear about where the model sits with respect to other, empirically-validated theory. There are three broad possibilities, but with no hard division between them. The model may:

- try to provide explanatory mechanisms for existing descriptive models, aiming to provide a more solid theoretical framework which might have more descriptive potential;
- explore different representations of existing theory, or the effects of different theoretical assumptions, with the intent of identifying problematic or underused theoretical avenues;
- explore the qualitative dynamics (primarily) and system level properties of some generalised abstraction of a real-world concept, normally with the intent of broadening theory or making parallels between normally distinct disciplines.

Only the third of these is what Windrum et al. call "synthetic artificial worlds which may or may not have a link with the world we observe" (Windrum *et al.* 2007, §4.3). The others are just 'traditional' models which are choosing to focus on analytical adequacy. (The second two are both using simulation for what Axelrod calls *discovery*: using the model "for the discovery of new relationships and principles" (Axelrod 1997).)

We have just seen an example of the first type: Sallans et al.'s agent-based model of integrated consumer and financial markets (2003). This was also used as a descriptive model, but only to replicate the empirical fit of other descriptive models (via stylised facts).

A fairly prototypical example of the second type would be the use of an individual-based model to compare with a system level one: the latter are "state variable models" in Grimm's terminology (Grimm 1999). The aim here is to show that the individual basis, together with some empirically-sound form of individual variation, significantly affects the dynamics in a way which suggests that the state variable model might be inappropriate in at least some circumstances. In this particular case, we can call this a "paradigmatic" motivation for the model (again, using Grimm's terminology): the comparision is between models with different ideological principles (and hence different positions on our dimensions here)[9]. Equally, there could be a less radical change to the model: to show, for example, that some mechanism originally omitted because it had no significant impact did, in fact, have an impact which had been overlooked by the original researchers.

Theoretical models of the third type are common in the field of ALife, since this is concerned (amongst other things) with trying to abstract the essential qualities of life and, by investigating synthetic alternatives which reproduce these qualities, to broaden theory that is restricted to life as it happened on Earth (Langton 1987). However, ALife also encompasses theoretical models of the *first* type (e.g., Hinton & Nowlan's demonstration of the Baldwin effect, as discussed by Di Paolo et al. (2000)) *and* descriptive usages which attempt to match empirical data and make predictions (again from Di Paolo et al. (2000), where they discuss the "virtual biology laboratories" of Kitano et al.). These latter models are classed as ALife because they are attempting to model biological systems as complex adaptive systems (CASs) via simulation (i.e., they are still using a computational modelling approach which is considered 'artificial' by mainstream biology).

[9]Of course, there are other, less paradigmatic motivations for using individual-based models (see the *Structural – Individualist* dimension). However, we are concerned here with cases where the model is being used as a theoretical model.

Because of debates regarding the scientific value of these differing approaches (e.g., strong versus weak ALife), ALife research has included considerable methodological discussion which is applicable here (Noble 1997; Di Paolo *et al.* 2000; Noble *et al.* 2000; Wheeler *et al.* 2002). Di Paolo et al.'s concept of **opaque thought experiments**, which came out of this debate, is a very useful description for *all* the types of theoretical model discussed here. Theoretical models:

- function as classic *thought experiments* in giving clarity and precision to the theoretical consequences of some postulated concepts (normally in a way which shows dissonance in current theory or a potential new direction);
- are *opaque* in that the complex consequences are not self-evident, and thus simulation is required— additionally, an understanding of how the model has produced the results that it did is needed, and it is often only this understanding which can be used to make an effective theoretical argument.

3.1.4 The Pseudo-Engineering Approach

We define a **pseudo-engineering approach** as one where a model which has had little or no theoretical validation is applied directly to the problem at hand as an applied 'engineering solution', *without achieving any significant predictive accuracy*. (As we have reiterated, this lack of predictive accuracy is typical for social systems simulation.)

The most common example is where the model is designed ad hoc to fit a particular problem. The results of such research are normally achieving some ontological adequacy (often via stylised facts) for current or historic data. In cases where the model provides some prediction of the real-world system's behaviour under conditions different from the current (e.g., under proposed new market rules), there may not even be this ontological adequacy, it being deemed that the system is different enough in the future situation that empirical validation against current or historic data is meaningless.

This is *pseudo*-engineering because, in reality, the theory is *not* valid to anywhere near the same degree as those of the natural sciences used in standard engineering: neither analytical/causal adequacy, nor indirectly via predictive accuracy. We stress that this is perfectly acceptable science (as discussed in the epistemological model), *but that the researcher should recognise what scientific value this limits them to.*

Firstly, because of the issues with model formalisation and empirical testing for social systems discussed earlier, there is much less certainty that a good empirical fit (without predictive accuracy) implies any kind of strong explanatory or predictive model.

Secondly, there are many competing paradigms in social science which are reflected in the taxonomy here. Thus, it is important for models to clearly contextualise themselves and reflect what theoretical conclusions they may or may not bring compared to other approaches to a general research question. That is, *exploration as a theoretical model can add significant scientific value*. Grimm is very insistent on this point in discussing how individual-based models in ecology are compared to the traditional, differential equation based theory of mathematical biology:

> "Individual-based modelling, on the other hand, would without reference to the conceptual framework of theoretical population ecology ultimately lead to mere 'stamp collecting', not to theory." (Grimm 1999, §4.6)

Thirdly, and more pragmatically, explicit analytical adequacy consideration helps avoid theoretical challenges to the work later or, worse, evidence that the bridging argument is unsound. (Read (1990) provides examples of how this can come unstuck.)

3.1.5 Stakeholder-Centric Approaches

We have seen that such approaches tend to reject the idea of a definitive real-world data generation mechanism and focus on adding precision to debate, not empirical accuracy. (Moss (2008) covers this in detail.) However, this does not mean that this *Theoretical – Descriptive* dimension is inapplicable to such approaches, which will typically be a mixture of theoretical and descriptive uses.

In terms of analytical adequacy, the focus is on the formal representation of subjective stakeholder views (subjective theory); ontological adequacy occurs via validation techniques such as Turing type tests, where the stakeholders agree that the outputs are consistent with their expectations. (This does not preclude more quantitative comparisons where the modeller and stakeholders feel it is relevant.) The specific ideological differentiators of stakeholder-centric approaches are captured more definitively in other dimensions.

3.1.6 Abductive Simulation and Bridging Arguments

The uniqueness of Werker & Brenner's abductive simulation approach (2004) relates to this dimension: it focuses on a *deferral of the specificity of the bridging argument between theoretical and descriptive uses* (refer back to figure 5).

It does this by including all empirically-supported theories in the initial model for a relatively broad class of real-world systems: say, a number of different countries' markets for X, where no a priori decision is made as to whether a particular country Y might require a particular mechanism or complete set of mechanisms, whilst others do not. A final abductive step is used to try to tease out the most meaningful classifications and relationships from sets of model configurations and empirical datasets which show a good empirical match. That is, which particular systems may be well represented by some variant of the generalised model is a decision which is deferred until after a full set of empirical comparisons is done; this initial process is therefore computationally and statistically intensive.

3.2 Dimensions Concerning Model Structure

These concern the nature of the mechanisms which will make up the model. As Eason et al. pithily put it (Eason *et al.* 1997), simulation models have aspects which are: supposed to correspond to reality, and are posited as making a difference; do not correspond to reality, but are posited as *not* making a difference. Of course, for any *particular* system (or class of systems), there will be other structural design decisions which are considered important, such as LeBaron's summary for agent-based market models (LeBaron 2001). However, we are concerned here with those affecting the scientific positioning of the model.

3.2.1 Structural – Individualist

To use McKelvey's terminology (McKelvey 2002), we can state this as a debate regarding how "idiosyncratic microstates" (heterogeneities amongst system participants performing the same roles) are treated. A structural approach focuses on the influence of social structures (e.g., institutions and firms) and the aggregated behaviour of individuals; thus, it assumes away or statistically treats individual variations. An individualist approach attempts to analyse the emergent structure, by explicitly modelling this variation. Thus, an individualist might attempt to explain an organisation's decision making as the outcome of individual behavioural differences and social interactions within the enterprise; a structuralist might explain it according to structural goals for the organisation as an aggregated entity (e.g., the profit maximisation view of neoclassical economics)[10].

[10]This dimension echoes one of the two used by Burrell & Morgan to characterise sociological paradigms (Burrell & Morgan 1979): namely the one contrasting theories emphasising *objective, structural* aspects of society with those emphasising *subjective, individualistic* ones. However, they imply a lot a philosophical baggage which is not always relevant here.

An individualist approach is often tied up with the notion that complex sets of *interactions* amongst (potentially differentiated) individuals may produce structural regularities which are not obvious a priori, without recourse to any centralised mechanism. There has been some success in reproducing specific patterns such as ant foraging and traffic jam formation (Resnick 1997), as well as more general fundamentals of human society (e.g., Epstein & Axtell (1996): this included tribal units, credit networks, and persistent social inequality). However, in our definition here, a structural approach may still be interested in the global effects of complex interactions, but potentially between more aggregate level entities, and without an emphasis on individual variation. The system dynamics approach (Sterman 2000, for example) is a good example of this, with its use of coupled differential equations and feedback loops representing the interaction of various aggregate processes. As an illustration, Bonabeau (2002) gives the example of a product adoption model which can be treated in a system dynamics or agent-based manner, with largely identical results (the system dynamics model reflecting the mean-value of the outcomes of individuals' interactions). It is only when the agent-based model (ABM) considers individuals *estimating* adoption rates from interactions in a spatial neighbourhood (rather than having global knowledge of them), that any significantly different dynamics arise.

Intermediate positions will typically reflect the observation that there tends to be a two-way process, with individual actions forming social structures which themselves influence or constrain individual action; an observation formalised in structuration theory, as discussed by Gilbert (1996).

We should note that an investigation often naturally follows a mixture of the two extremes (rather than a single extreme or intermediate position), depending on its focus. Take Ladley and Bullock's study of the effects of interaction topologies (market segregation) on market convergence (Ladley & Bullock 2007). This is individualistic in terms of being agent-based, with adaptive learning algorithms which mean that individuals will develop idiosyncratic behaviour[11]; yet, it is structural in imposing abstracted interaction topologies a priori.

Modelling Techniques. Whilst we need to be aware that modelling techniques are just *tools*, not ideological statements in themselves, some lend themselves more readily to some positions than others. In the case of this dimension, approaches such as agent-based models and cellular automata (CA) are clearly a strong fit for an individualist approach.

[11] In this case, there is actually no individual variation in the algorithm or its parameters, but random terms ensure that learning differs per individual.

Agent-based models which arise from a strong individualist position can be regarded as paradigmatically motivated, and will tend to refer back to structural theory in an attempt to show how the individualist approach shows real-world effects which are more difficult to achieve with structural approaches. Others may use ABM just for its fit to the problem at hand, or its naturalness of representation: "pragmatic motivation", to use Grimm's phrase (Grimm 1999).

Micro-Realism and Macro-Realism. Not all individualistic approaches are valid: there is one particular methodological error that is worth mentioning because we believe that it is more prevalent than one might think, and because it relates to other dimensions in our taxonomy.

One of the appeals of agent-based models is that they allow a more natural representation for most social systems, since there are explicitly modelled individuals and the designer can code behaviour as a set of decision rules which match more cleanly to participants' or observers' verbal descriptions of how choices are made. Well presented models of this type, particularly the influential ones of Epstein & Axtell (1996), demonstrate that such a 'realistic' approach can generate dynamics which match those in real-world systems (at least qualitatively), and that the emergence of such global behaviour from decentralised local rules may apply to a wide variety of social systems.

This is all fine as it stands, but problems occur when modellers take this too uncritically as a belief that there is some kind of methodological magic in ABM which *'automatically' causes realism at the micro level to be translated into realism at the macro level of system level patterns*. We call this the **micro-realism implies macro-realism (MiRIMaR) fallacy**. Grimm captures the idea well:

> "Kaiser (1979) comments that individual-based modelling 'is naive in the sense that it directly relies on observed data and interrelations' (p.134). This naivety bears the risk that modelling is no longer regarded as a mental activity, but as something that is done by the model entities themselves: simply cram everything you know into a model and the answers to the question at hand will emerge via self-organisation. But this never happens." (Grimm 1999, §4.1)

This tends to manifest itself in the choice of validation conducted (or, rather, not conducted). It is also closely related to some other problematic steps:

- premature inference that a theoretical model can also act as a descriptive model (*Theoretical – Descriptive* dimension);
- an uncritical choice of the physical individual as the only correct aggregation level for the model (*Simplifying Refinement – Additive Refinement* dimension).

3.2.2 Apriorist – Empirical

The essential distinction here is how much of the model's design should be based on hard, empirical data, as opposed to a priori assumptions—what Brenner & Werker call:

> "assumptions [. . .] based on theoretical considerations, which result from axioms, ad-hoc modelling or stylised facts." (Brenner & Werker 2007, §3.2)

Rejection of Abstraction via Apriorism. Where apriorism is explicitly rejected, there is a belief that models should *only* include causal mechanisms which have been seen empirically. As an example, consider Brenner and Murmann's simulation of the synthetic dye industry (Brenner & Murmann 2003). To be valid in their eyes, their inclusion of chemist migrations to other countries as a modelled process had to have confirmation from documentary evidence that this happened (and presumably in large enough numbers, or with a small enough total number of chemists, so that it could be justified as representing a potentially significant process). Epistemologically, this is therefore focusing heavily on causal adequacy.

This position is particularly against strong apriorist assumptions which are exempted from empirical validation, such as neoclassical economic assumptions of rationality and general equilibrium. It is also against weak apriorism (where the assumptions *are* subjected to empirical validation). We can see the latter from Brenner and Werker's quote above, where they reject theory based on stylised facts (which we have already defined as empirically-backed patterns demonstrated via a previous descriptive model). Despite being empirically backed, they presumably see such patterns as merely heuristics or selective statistical matches, and value only actual quantitative values (e.g., total number of firms or the market share of a particular country's firms, as used in their dye industry model).

As a counterpoint to this view, the epistemological model showed that the process of observing and selecting empirical data is inextricably linked with theoretical issues. Therefore, we can never be totally free of a certain spectre of apriorism— a point conceded by Brenner and Werker (2007, §3.1).

Use of all Available Socio-Historical Data. There is the related question of what types of empirical data to consider.

Social science has a long tradition of using more sociohistorical oriented approaches. Amongst other things, this supports a focus in social systems analysis on how involved individuals perceive the system, and thus on testimonies, case studies and the like. Windrum et al. (2007) discuss case-study oriented models (the "history-friendly approach"), whilst Edmonds and Moss (2005) put forward the strong position that *all* potential types of empirical evidence (including anecdotal evidence) should be considered:

> "[...] if one has access to a direct or expert 'common-sense' account of a particular social or other agent-based system, then one needs to justify a model that ignores this solely on *a priori* grounds."

They stress that simulation (particularly agent-based) and modern computing power provide the opportunity, not available in an analytic approach, to include all such ideas and then weed out inappropriate ones via many simulation runs and analyses.

As well as focusing on the empirical *evidence* used to build theory, this stance also tends to endorse a stakeholder-centric modelling process. The epistemological relativism of such approaches is part of the separate *Universally Simple – Naturally Complex* dimension.

3.2.3 Universally Simple – Naturally Complex

Similarly to the previous dimension, this has two closely intertwined threads: whether simplicity is a valid criterion of adequacy, and whether models can be more universal or only applicable to very specific contexts. Typically, approaches favouring simpler models see them as more universally applicable, but it *is* possible to have 'inconsistent' stances on these two areas, as we will discuss (e.g., favouring simple but very specific, non-universal models). However, this is rare enough (and problem specific enough) that this does not merit splitting this into two dimensions.

Rejection of Simplicity as a Driver for Abstraction. Some academics reject the basic idea of simplicity in a model being a good criterion for asserting the validity of one social model over another, despite this being a strongly held criteria of adequacy for the physical sciences; this is effectively a rejection of Occam's razor.

This relates to the fundamental perceived complexity of social systems. To quote Edmonds and Moss from their KIDS approach:

> "[This] domain of interacting systems of flexible and autonomous actors or agents [means that] the burden of proof is on those who insist that it is not sensible to try and match the complexity of the model with the complexity of the phenomena being modelled" (Edmonds & Moss 2005, §2).

Thus, though complexity theory shows that complexity *can* arise from algebraically simple mathematics, this is not a licence to apply simplicity criteria indiscriminately. Equally, though simplicity might aid model analysis, this may be largely irrelevant if some irreducible level of complexity is needed for descriptive accuracy (and, if accurate, model simplifications could always be sought *later*— see the *Simplifying Refinement – Additive Refinement* dimension).

The main counter to this is as expressed by Grimm (1999): modellers should "adopt the attitude of experimenters", so as to understand *how* the model produces the results that it does. Treating models as black boxes means that analytical adequacy is going to be limited to results-based comparisons with other theory, so this criticism is more relevant to models which are not solely acting as descriptive models. This argument is diluted slightly by the steady increases in computational power, which mean that a complete sensitivity analysis can be performed on even quite complicated models in a realistic time frame (though this is no substitute for true understanding).

Universality and Relativism. It is not necessarily true that a more simple model will be more universally applicable (generalisable) than a more detailed one, so universality versus relativism becomes a separate issue. A simple, abstract model could provide predictive accuracy only for a very specific system: perhaps for some highly specialised, predator-free animal species, whose behaviour was dominated by a small number of generalised physiological, psychological and/or environmental characteristics (so an abstract model would suffice, but would not be generalisable to other species which existed in more complex ecosystems).

Sociohistorical approaches tends to reject universality and promote a more contextual approach, where community-specific cultural factors have more impact on behavioural patterns than biological traits. This goes hand-in-hand with another form of contextualism: the assertion that there will generally need to be *multiple* models to explain behaviour, since different system-level effects may be due to very different causes at different levels of aggregation and at different points on this biological–cultural axis. This fits with our epistemological model, particularly Azevedo's map analogy. Different maps (models) accentuate different features of the 'landscape' and are suitable for different end uses. Such a division reflects the realisation in sociology that competing schools of thought could sensibly work together, as noted by Eisenstadt & Curelaru in the 1970s (Eisenstadt & Curelaru 1976, p.372).

Where simulations are used to support decision-making, this contextualism can manifest itself as a stakeholder-centric approach, as discussed previously. The companion modelling approach exemplifies a more extreme, relativistic stance; one where the possibility of prediction beyond the short-term is rejected entirely, and the emphasis is switched to the subjective understanding of the system by its stakeholders. Moss makes this clear (Moss 2008), asserting that "forecasting over periods long enough to include volatile episodes cannot be reliable and, as far as I know, has *never* been observed", stressing that this implies that "The purpose of the models themselves is to introduce precision into policy and strategy discussions".

Such a view is much more mainstream in policy-related work not based on simulation, such as spatial decision support systems (SDSSs). The primary objective there is normally to provide a more objective and precise view of conflicting stakeholder interests (e.g., wind farm owners, countryside groups and local residents for a wind farm siting problem). A spatial model allows the implications of different weightings of interests to be observed (in terms of their constraining effect on potential sites), aiding in decision making (see Carver (2003) for a summary, albeit in the context of the general public as the stakeholders in environmental decisions). In *simulation* work, however, we are concerned with playing out the consequences of stakeholder ideas on the causal processes underlying a system's behaviour.

Moss goes on to argue that this means that, since the models are contextual to the participants, there is no epistemological value in saying that one model is theoretically better than another if they both model the empirical data equally accurately. That is, some universal brand of analytical adequacy is inapplicable, since the theory is subjective.

3.3 Dimensions Concerning System Exploration

These relate to how the model is intended to be used in exploring the dynamics of the system in question.

3.3.1 Stability – Radical Change

This dimension contrasts theories emphasising regulation and stability versus those emphasising radical change[12]. Moss (2008) characterises the debate well. A radical change oriented approach is primarily based on the "unpredictable, episodic volatility" of most social systems, and evidence that these episodes typically alter the structure of the system in fundamental ways. A stability oriented view will be more interested in equilibria and asymptotic behaviour (an approach often associated with neoclassical economics), implying a belief that, even if state changes occur, they are infrequent, predictable or avoidable enough that a stability-oriented exploration can meaningfully explain, characterise or predict general behaviour. Models which rely on historical data to predict future values[13] can also be viewed as strongly stability-oriented approaches, in that they assume that there is some structurally stable pattern over time.

Note that a radical change approach does not imply a rejection of continuous, differential equation based techniques (such as system dynamics). Chaos theory has demonstrated that even algebraically simple such systems can show extreme volatility and unpredictability. In addition, the combination of such equations in feedback loops (in system dynamics) can produce varying behaviour, from stable equilibria to periodicity to chaos. A stability oriented approach would focus on the parameter ranges where stable behaviour occurred (and their relationships to empirical values), whereas a radical change one might choose to emphasise that volatile behaviour occurred at commonly occurring values.

As Richiardi et al. point out (2006, §4.8), individual-based models need to be clear about whether they are considering micro or macro-level equilibria in a stability focused approach.

3.3.2 Qualitative Validation – Quantitative Validation

This dimension is probably the one most influenced by other dimensions. For example, models taking a more abstract, structural approach will tend to favour validation via qualitative features of the real-world system, such as stylised facts. Models based more on context-specific, empirically-aligned theory will tend to focus on specific validation against quantitative global values (e.g., market prices); approaches such as Bayesian simulation

[12] As for the *Structural – Individualist* dimension, this also echoes terminology used by Burrell & Morgan in characterising sociological paradigms (Burrell & Morgan 1979); however, this time, the meaning here corresponds fairly well with theirs.

[13] Such as Marmiroli et al.'s stochastic model of an electricity market (2007), which provides risk analysis via Monte Carlo simulation. Factors used in the risk model are represented by econometric models based on historic data.

are wholly centred around a quantitative, statistical philosophy for validation and parameter space exploration (Brenner & Werker 2007).

However, this is not always such a direct correlation. Forecasting models, despite often being abstract and structural (e.g., an econometrics approach), sometimes rely on quantitative validation against known time series of particular variables. In addition, differing *scenarios* can be used as a qualitative approach to complement quantitative ones in looking at the robustness of the model. These can be in two forms:

Theory-Aligned Abstractions. The use of abstracted scenarios (e.g., perfect market competition) to confirm that the model aligns with more classical theory in 'limiting cases'; Grimm (1999) calls these scenarios "strong cues". We see this technique in Epstein and Axtell's agent-based Sugarscape models, where they run a scenario with infinitely lived agents and fixed trading preferences, so as to mimic the assumptions of neoclassical economics (Epstein & Axtell 1996, §4).

As Grimm points out (1999), strong cues are also invaluable in attempting to understand the mechanisms at work inside the (black-box) model, since they potentially represent strong boundary conditions and simpler parameter sets which may not include the 'noise' of a fully empirically calibrated one. In a nutshell, this technique represents a *theoretically-guided exploration of the parameter space*, which is analogous to Richiardi et al.'s "global investigation" (2006, §4.14).

Historical Alternatives. The use of alternative scenarios to investigate possible other outcomes and the different qualitative effects[14]. However, this bears the danger that, given sufficiently different scenarios (from some base case), our assumptions about the system structure and mechanisms may no longer apply there (i.e., we might not be able to capture the full range of behaviour in a single model— or at least not at the same level of aggregation— so our comparison becomes meaningless).

This is particularly acute for comparisons of models against historical data, where we know the actual behaviour which occurred. Windrum et al. (2007) use the term "counterfactual histories" in such cases, and point out potential difficulties where, say, we are modelling a market where the actual history depended crucially on a particular action from a particular firm. In this case, tests using parameter values other than the 'real' ones may be of questionable use,

[14] We might also compare *quantitatively*, particularly if the scenarios were fairly similar to each other.

since the system might have had a radically different structure in such cases; or, similarly, our model structure may have been overly influenced by the actual history when, in fact, this represented a very unlikely outcome of the 'true' model.

3.3.3 Simplifying Refinement – Additive Refinement

This covers how the model is refined during the course of the research. Does it: start detailed and then get progressively simplified via sensitivity analysis or similar? start at a coarse level and get made more complex until the desired empirical accuracy is reached? take a more mixed approach or not consider refinement? In addition, is the refinement being done to move towards a desired aggregation level, or is the refinement just to make the model more accurate or simple (i.e., already starting at the desired aggregation level)?

Edmonds and Moss's KIDS approach (Edmonds & Moss 2005) starts with a model detailed enough to include mechanisms based on all available empirical evidence. Refinement is guided by what "evidence and [increased] understanding of the model support", and may be additive *or* simplifying.

Other approaches are more intrinsicially wedded to one style of refinement. Grimm (1996), for example, advocates a model which starts at a high level of aggregation, and uses validation against some pattern in the empirical data to guide its refinement towards the desired aggregation level (as required to address the research problem in question). Because the pattern relates to a specific spatial and/or temporal scale, it can be used to direct and constrain each refinement in aggregation (because the pattern needs to be maintained at its appropriate scale).

Brenner and Werker's abductive simulation approach (Brenner & Werker 2007; Werker & Brenner 2004) begins with the same kind of 'sufficiently detailed' model as the KIDS approach. However, the emphasis is then on extensive use of empirical data to simplify the model, in terms of constraining parameter values and rejecting model alternatives which do not match the data (all over a set of empirical instances for similar types of system, which facilitates the later abductive step which is the main differentiator of their approach).

4 Conclusions

We have presented a taxonomy which attempts to define the most crucial dimensions for the methodological positioning of social simulations. This is backed up by an overarching epistemological model, which provides the context in terms of what knowledge the model is trying to achieve (and introduces important concepts such as theoretical and descriptive usage). The main aims are to:

- Provide all modellers with a framework to help understand how they and others make decisions on the scientific positioning of their research, especially in understanding the often implicit positioning choices made by their particular discipline. We hope that this can also be a step towards more useful interdisciplinary discussion, and perhaps a shared vocabulary.

- Provide social simulation 'methodologists' with a useful and philosophically 'sympathetic' summary of scientific positioning concerns, as well as to stress that these are fundamentally intertwined with other methodological choices. The hope is that this can stimulate further discussion on improved protocols and standards for social simulation.

Below, we briefly compare our taxonomy with some others in the literature, before discussing the current limitations and potential further work to progress the aims above.

4.1 Comparisons with Other Taxonomies

Brenner & Werker (2007) propose a two-dimensional taxonomy concerned with the inference process by which models try to achieve scientifically credible results:

- hypothetical–empirical;

- general–specific.

The former covers the degree to which empirical data is used in model formulation, and hence maps roughly to our *Apriorist – Empirical* dimension. The latter covers how many model specifications are investigated: "In specific analyses only one specification of the simulation model is extensively simulated and the results analysed." (Brenner & Werker 2007, §3.1). This is really a categorisation of *what* validation and sensitivity analysis is done, not really the ideological *why* behind it. From our point of view, this dimension would really be a consequence of decisions made along all our dimensions.

The authors also see the validation process as *changing* the positioning of models along these axes, since increased use of empirical data in validation makes the model both less hypothetical and less general (since it is being compared with specific empirical data sets). This is problematic since, unless they are including causal adequacy tests *outside* the scope of the simulation process, they are conflating model formulation with model parameter constraint: the hypothetical (apriorist) nature of the mechanisms is not affected by how tightly constrained the model parameters are by empirical validation. As we have seen, descriptive accuracy does not necessarily imply plausible (or 'empirical') explanatory mechanisms.

Moss (2008) uses a related single-dimension taxonomy: "the metaphor of a spectrum of models ranging from the most theory-driven to the most evidence-driven". This is effectively the same angle as that taken by Brenner and Werker, albeit expressed in a different way and for a different purpose. Both are thus data-oriented characterisations.

Finally, social theorists such as Burrell & Morgan (1979) and Eisenstadt & Curelaru (1976) have attempted to characterise the paradigms of sociological theory, and their development over time. These analyses certainly help clarify the general sociological schools of thought, but the aim here is to pragmatically characterise social *simulations* in particular and the dimensions which most govern the model design and investigation process adopted.

4.2 Limitations & Further Work

Firstly, there is no real discussion on possible trade-offs between decisions made on different axes. We have attempted to make them as independent as possible, but we should bear in mind the possible validity of arguments such as Levin's influential one in population biology (Levins 1966): that generality, realism and precision cannot be satisfied together in any model; one must always be sacrificed.

Secondly, there need to be further attempts to use the framework to characterise debates and disciplinary differences in the field. Appendix A supplements the main text in this regard, but feedback from other academics is really required. For example, take the following selection from Swedberg et al.'s summary of the differences between economic sociology and neoclassical economics (respectively) as approaches (1987):

- the actor as a social actor vs. a separate utility maximiser;
- actors' actions as social rationality vs. formal rationality;

- the results of economic action as tension-filled interest struggles vs. equilibriated harmony;
- the general scientific method as descriptions and explanations based on empirically adjusted abstractions vs. predictions and explanations based on radical abstractions.

We can see direct correspondences to our epistemological model and taxonomy, but it would be very useful to get the opinions of researchers directly involved in these disciplines and their 'definition'.

Thirdly, the framework focuses only on epistemological and methodological aspects of scientific positioning. It cannot therefore help answer all modelling questions, such as (assuming an agent-based model), 'What should my agents represent?'. In this case, a decision on the *Structural – Individualist* axis will give *some* idea, but domain specific knowledge will always be needed to make the final choice, which may also have to consider other factors, such as: researchers' individual experience and skillsets, the nature of the empirical data, and the computational resources available.

It is not therefore 'integrated' with more descriptive protocols (or social simulation textbooks), but we feel that there needs to be feedback and considerable further discussion with such methodologists before this is taken forwards. In particular, there is the key question of how much a 'typical' modeller *needs* to understand positioning issues to the depth given here, and to what degree they should spell out their decisions in papers.

On the one hand, we believe that such knowledge is important, precisely because there are many rival approaches, and thus a sea of often confusing terms and concepts. In this sense, social modelling is seeing interesting times as, for example, agent-based 'bottom-up' approaches challenge the orthodoxy of microeconomic theory. If we use Kuhn's view of science again (Kuhn 1970), there are many aspects of a crisis period in the science, which makes it crucially important to have clear dialogue on epistemological and methodological issues (even if this ends up being an awareness of each other's positions, without compromise on either side).

On the other hand, most scientists doing 'normal' research stay within the bounds of their particular disciplinary paradigm, and this is what is required to progress science without constant questioning:

> "Normal scientists must be uncritical of the paradigm in which they work. It is only by being so that they are able to concentrate their efforts on the detailed articulation of the paradigm and to perform the esoteric work necessary to probe nature in depth. [. . .] Much of the normal scientist's knowl-

> edge [of the precise nature of their paradigm] will be *tacit* [. . .]" (Chalmers 1999, talking about Kuhn's theory)

So perhaps such discussion should be left to a relatively small group of methodologists, rather than being an explicit requirement for all.

5 Acknowledgements

This research is part of work supported by the UK EPSRC research council in CASE studentship CASE/CNA/07/56, which is in collaboration with National Grid PLC. We thank the two reviewers of this work, particularly Volker Grimm who 'unanonymised' himself in order to give more directed feedback.

Appendices

A Appendix A: A Summary of Dimensional Classifications for some Key Approaches

The summary here should help draw together the arguments elsewhere in this paper. Table 2 shows a compact view of different approaches' positioning along the various dimensions.

The abbreviations used for the names of dimensions are as below:

T–D Theoretical – Descriptive.

Model Structure Dimensions

S–I Structural – Individualist.

A–E Apriorist – Empirical.

US–NC Universally Simple – Naturally Complex.

System Exploration Dimensions

S–R Stability – Radical Change.

Ql–Qn Qualitative Validation – Quantitative Validation.

SR–AR Simplifying Refinement – Additive Refinement.

Approach	Dimensions						
			Structural			*System Exploration*	
	T-D	**S-I**	**A-E**	**US-NC**	**S-R**	**Ql-Qn**	**SR-AR**
Neoclassical economics	-	S	A	US	S	(Ql)	AR
Agent-based computational economics (ACE) e.g., Sallans et al.	(T)	I	(A)	(US)	-	(Ql)	(AR)
Companion modelling	-	-	E	NC	R	-	(SR)
History-friendly and abductive simulation	(D)	(I)	E	(NC)	-	Qn	SR
KIDS (Edmonds & Moss)	-	(I)	E	(NC)	-	-	(SR)
Grimm's ecological modelling recommendations	-	I	(A)	(US)	(S)	Ql	AR
ALife as opaque thought experiments (Di Paolo et al.)	T	I	A	US	-	Ql	-

Table 2: A summary of dimensional positioning for various approaches. Brackets indicate a weak correlation with the given alternative. A dash indicates a neutral or balanced position towards the dimension

Clearly, table 2 omits some of the subtleties, but it helps highlight how the taxonomy provides useful, 'sharp' differentiations between approaches. A slightly more detailed explanation of the table contents follows (though still with some unavoidable generalisations).

Neoclassical economics	
Theoretical – Descriptive	Neutral. Concerned with theoretical issues, but extensively used to try to characterise and predict macroeconomic trends.

Structural – Individualist	Strongly structural, concerned with system level properties and equilibrium.
Apriorist – Empirical	Strongly apriorist. A priori assumptions and idealised mechanisms.
Universally Simple – Naturally Complex	Universally simple. Tends to model consumers, firms, etc. in general, without distinctions for, say, specific markets.
Stability – Radical Change	Stability. Almost wholly concerned with equilibria and how markets find/return to them.
Qualitative Validation – Quantitative Validation	Tends to be qualitative (fit to stylised facts), but also used for quantitative macroeconomic prediction.
Simplifying Refinement – Additive Refinement	Additive. Start with simple models and refine if necessary (honours the criteria of adequacy in the physical sciences).

Agent-based computational economics (ACE)

Theoretical – Descriptive	In theory, neutral as for neoclassical economics but, in practice (at least currently), is often concerned with showing inadequacies with the neoclassical approach and so can perhaps be better characterised as weakly theoretical.
Structural – Individualist	Individualist. By definition of agent-based approach.

Apriorist – Empirical	Weakly apriorist. Still takes some neoclassical (abstract) theory as starting point but tends to add more empirically observed behaviour.
Universally Simple – Naturally Complex	Not as universal as neoclassical economics, but still tends to look at general cases (despite the implied complexity of the agent-based approach).
Stability – Radical Change	Tends to be neutral. ABMs can show stable or radical system level effects— the approach tends to see which emerges and treat this as one of the main areas of interest.
Qualitative Validation – Quantitative Validation	Weakly qualitative, as for neoclassical economics.
Simplifying Refinement – Additive Refinement	Tends to be additive, as for the neoclassical approach, adding different individual variation to see how this changes the behaviour, although this is not a hard and fast rule.
Companion modelling	
Theoretical – Descriptive	Balanced. Validates against both subjective stakeholder theory and aims for descriptive accuracy (although to stakeholder-specific criteria, such as Turing style tests).
Structural – Individualist	Neutral. The approach will depend on the modeller and stakeholders.
Apriorist – Empirical	Empirical. Attempts to capture all potential mechanisms from stakeholder anecdotal evidence, etc. Rejects apriorism.

Universally Simple – Naturally Complex	Naturally complex. Strongly relativistic position that models may only be valid to individual stakeholders.
Stability – Radical Change	Rejects predictive accuracy because of unpredictable volatility of social systems, so a strong radical change view. However, in the scope of a particular model (valid only to make precise stakeholder ideas and their short-term implications), model analysis is neutral in this regard.
Qualitative Validation – Quantitative Validation	Neutral. The stance will tend to depend on the modeller and stakeholders.
Simplifying Refinement – Additive Refinement	Tends to be simplifying since the initial model incorporates all possible mechanisms empirically identified, but not always (dependent on stakeholder perceptions).
History-friendly & abductive simulation	
Theoretical – Descriptive	Weakly descriptive. Tends to focus on descriptive models due to context-specific nature of sociohistorical view.
Structural – Individualist	Will tend to model at the aggregation best aligned to case studies, which will often reflect the important role of specific firms, individuals, etc. Therefore, weakly individualist in approach.
Apriorist – Empirical	Empirical. Rejects apriorism and focuses on numerous empirical sources.
Universally Simple – Naturally Complex	Weakly relativistic. Focuses on context specific elements identified from case studies or similar but not as strong a stance as companion modelling.

Stability – Radical Change	Neutral. Will depend on the particular historical context. Possibly leans slighty towards radical change due to the historical view of social struggle.
Qualitative Validation – Quantitative Validation	Quantitative. Focuses on extensive validation against empirical data and, particularly for abductive simulation, focuses on figures from case studies.
Simplifying Refinement – Additive Refinement	Simplifying. Begins with model incorporating all potential mechanisms and simplify as appropriate from empirical comparison.
KIDS (Edmonds & Moss)	
Theoretical – Descriptive	Neutral. Focuses on descriptive models but the KIDS approach also allows for the analysis of several of these to look for more abstract general theory.
Structural – Individualist	Weakly individualist. Focuses on the aggregation level that varied empirical data supports, but this will tend to to focus on variation amongst 'individuals' (e.g., households in their UK water demand model).
Apriorist – Empirical	Empirical. Strongly rejects apriorism.
Universally Simple – Naturally Complex	Weakly relativistic and quite strongly rejects simplicity as a criteria of adequacy. Tends to focus on a contextual understanding of the system, similar to the history-friendly approach.
Stability – Radical Change	Neutral. No specific stance.

Qualitative Validation – Quantitative Validation	Neutral. No particular validation methodology.
Simplifying Refinement – Additive Refinement	Weakly simplifying, in terms of assessing descriptive models and looking for simplified abstractions. However, the basic descriptive models may be refined either way from the initial model.
Grimm's ecological modelling recommendations	
Theoretical – Descriptive	Neutral. Focus on descriptive accuracy (against patterns in the data) but also theoretical positioning and comparison with top-down approaches.
Structural – Individualist	Individualist. Specific use of IBMs (though advocates always considering structural properties so as to compare with top-down theory).
Apriorist – Empirical	Weakly apriorist. Focuses on comparison with top-down theory and hence on keeping a more abstract base but adding in individual variation.
Universally Simple – Naturally Complex	Tends towards universal, abstract models for the same reasons as above.
Stability – Radical Change	Weakly favours stability since focuses on looking at system level (emergent) properties and (stable) patterns in empirical data.
Qualitative Validation – Quantitative Validation	Qualitative. Advocates pattern-oriented modelling as best way to get more focused research results.

Simplifying Refinement – Additive Refinement	Additive. This ties in with the pattern-oriented method and comparison with top-down methods. There are also pragmatic reasons for avoiding simplifying refinement (no incentive for modeller to do it).

ALife as opaque thought experiments (Di Paolo et al.)

Theoretical – Descriptive	Theoretical. Concerned with thought experiments exploring the consequences of different theoretical assumptions and how this questions existing (structural) theory.
Structural – Individualist	Individualist due to ALife's links with complexity science and thus agent-based modelling.
Apriorist – Empirical	Apriorist. Concerned with fundamental abstractions of life-like phenomena.
Universally Simple – Naturally Complex	Universally simple, for same reasons as above.
Stability – Radical Change	Neutral. Systems may produce either behaviour, although there is normally an assumption that radical changes may well occur due to the system complexity.
Qualitative Validation – Quantitative Validation	Qualitative. Concerned with the overall qualitative system dynamics and how these compare to existing structural theory and more real-world instances of the system abstraction.
Simplifying Refinement – Additive Refinement	Neutral. May either tend towards additive (as for ACE) or, when looking for the most essential properties of life-like systems, take a simplifying approach.

B Appendix B: The Epistemological Model as an Improved Validation Framework

In terms of its validation terminology, our epistemological model can be directly compared to other validation frameworks (Stanislaw 1986; Bailey 1988).

B.1 Stanislaw's Validation Types

Stanislaw (1986) introduces his own set of terminology which attempts to isolate the components of overall validity. (These definitions are used by Richiardi *et al.* (2006, §4.28) in their methodological protocol.) He defines three canonical types of validation: theory, model and program validity. The last two correspond to our concepts of analytical and software adequacy (respectively). The other, theory validity, is defined as the validity of the theory relative to the simuland (real-world system). Rather than our more restrictive definition of causal adequacy, theory validity covers the general validity of the theory *as operating through the medium of the simulation*; i.e., a combination of ontological, software and analytical adequacy implies Stanislaw's theory validity[15].

This comparison is summarised in figure 11.

Looking at Stanislaw's clarifying example for expert systems provides some further insight. He argues that, since expert systems only attempt to *functionally* replicate the input/output processing of a target system (rather than modelling the 'true' internal processes), theory validity does not therefore apply (since there is no theory attempting to represent the true data-generation mechanisms of the real-world system). However, empirical tests of an expert system simulation versus the real-world system still test the simulation–phenomena link (as well as Stanislaw's model and program validity), but he leaves no term with which to describe this 'extra' validation, and can only use terms outside his definitions, such as the "overall" validity used by Richiardi *et al.* (2006, §4.29). It would seem much more natural to define a term like ontological adequacy to explicitly represent the simulation–phenomena validation.

This example also clarifies that what Stanislaw means by 'theory' is *explanatory* theory (as in figure 11). Expert systems still have their own AI-related theory, which can be explored via model validity, but this theory cannot be compared to the real-world system in question as a potential explanation. Despite the fact that Stanislaw states that the expert system

[15] This could arguably include causal adequacy as well, if we take the view that Stanislaw's theory validity also covers this 'extra-modelling' validity. However, this choice does not affect the argument in this section.

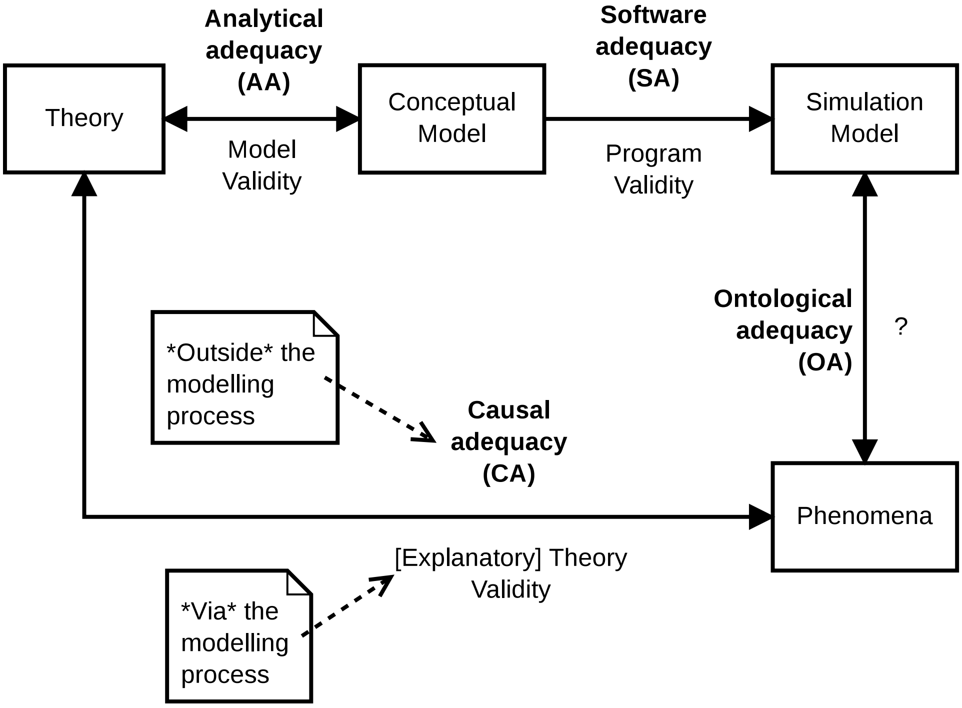

Figure 11: A comparison of Stanislaw's validation terminology (normal face) with those in this paper's framework (bold face). Note the lack of equivalent for ontological adequacy

aims to replicate the functionality of the real-world system, he seemingly does not consider that a potential *functional explanation* of it might be a useful explanatory result in itself. This is at odds with philosophers of science such as Grüne-Yanoff, who argues precisely this as a more appropriate aim for social simulations: "It [the simulation] suggests an analogy between the organisational structure of the simulator and the real-world system" (Grüne-Yanoff 2009, §4).

To use his example, a particular climate model used an "instability-dampener" component, despite knowing that there was no explicit, single real-world mechanism known to perform this function. By successfully replicating empirical data, this suggested that this set of functionalities (rather than specific real-world mechanisms) may reflect the organisational principles of the real-world system. That is, the theory may have been missing or misrepresenting some mechanisms which, together, produce some system-level instability-dampening behaviour in some unspecified way. This may mean that we should extend our causal adequacy to include some form of 'functional adequacy testing' as well, though this is not addressed further here.

B.2 Bailey's Operational and Empirical Validity

Richiardi et al. augment Stanislaw's list with operational and empirical validity (Richiardi *et al.* 2006, §4.29), as discussed by Bailey (1988). These cover the validation of the *indicators* used to formalise 'fuzzy' social concepts in their relationship to both the concept (Bailey's Type B validity), and the empirical reality (Bailey's Type C validity).

We regard these types of validation as already included within our existing set of definitions. Analytical adequacy covers the formalisation of the theory into a computationally implementable model (and thus Type B validity); causal adequacy covers empirical tests for the validity of the individual mechanisms (Type C validity, and possibly any empirical tests used to help support Type B validity arguments).

References[16]

AXELROD, R. (1997). Advancing the art of simulation in the social sciences. In: *Simulating Social Phenomena* (CONTE, R., HEGSELMANN, R. & TERNA, P., eds.), vol. 456 of *Lecture Notes in Economics and Mathematical Systems*. Springer, pp. 21–40.

AZEVEDO, J. (2002). Updating organizational epistemology. In: *The Blackwell Companion to Organizations* (BAUM, J., ed.), chap. 31. Blackwell Business, pp. 715–732.

BAILEY, K. (1988). The conceptualization of validity: Current perspectives. *Social Science Research* **17**, 117–136.

BARRETEAU, O. *et al.* (2003). Our companion modelling approach. *Journal of Artificial Societies & Social Simulation* **6**(2), 1. URL `http://jasss.soc.surrey.ac.uk/6/2/1.html`.

BONABEAU, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences* **99**, 7280–7287.

BRENNER, T. & MURMANN, J. (2003). The use of simulations in developing robust knowledge about causal processes: Methodological considerations and an application to industrial evolution. Papers on Economics & Evolution 0303, Max Planck Institute, Jena.

BRENNER, T. & WERKER, C. (2007). A taxonomy of inference in simulation models. *Computational Economics* **30**, 227–244.

BURRELL, G. & MORGAN, G. (1979). *Sociological paradigms and organisational analysis : elements of the sociology of corporate life*. Heinemann.

CARVER, S. (2003). The future of participatory approaches using geographic information: developing a research agenda for the 21st century. *Urban and Regional Information Systems Association (URISA) Journal* **15**(APA 1), 61–71. URL `http://www.urisa.org/node/806`.

CHALMERS, A. (1999). *What is this thing called science?* Open University Press, 3rd ed.

[16]Due to a quirk of the LaTeX packages used to prepare this post-print version, these references include an entry for the paper itself! Needless to say, this is not in the published version.

Di Paolo, E. A., Noble, J. & Bullock, S. (2000). Simulation models as opaque thought experiments. In: *Artificial Life VII: Proceedings of the Seventh International Conference on Artificial Life* (Bedau, M., McCaskill, J., Packard, N. & Rasmussen, S., eds.). MIT Press.

Eason, R., Rosenberger, R., Kokalis, T., Selinger, E. & Grim, P. (1997). What kind of science is simulation? *Journal of Experimental & Theoretical Artificial Intelligence* **19**(1), 19–28.

Edmonds, B. & Moss, S. (2005). From KISS to KIDS - an 'anti-simplistic' modelling approach. In: *Multi-Agent and Multi-Agent-Based Simulation (Joint Workshop MABS 2004)* (Davidsson, P., Logan, B. & Takadama, K., eds.), vol. 3415 of *LNCS*. Springer.

Eisenstadt, S. & Curelaru, M. (1976). *The Form of Sociology– Paradigms and Crises*. John Wiley & Sons.

Epstein, J. & Axtell, R. (1996). *Growing artificial societies: social science from the bottom up*. Brookings Institution Press.

Gilbert, N. (1996). Holism, individualism and emergent properties. an approach from the perspective of simulation. In: *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View*. Springer.

Gilbert, N. (2004). Quality, quantity and the third way. In: *Methods in development research: Combining qualitative and quantitative approaches* (Holland, J. & Campbell, J., eds.), chap. 14. ITDG.

Gilbert, N. & Troitzsch, K. (2005). *Simulation for the Social Scientist*. Open University Press, 2nd ed.

Goldspink, C. (2002). Methodological implications of complex systems approaches to sociality: simulation as a foundation for knowledge. *Journal of Artificial Societies & Social Simulation* **5**(1), 3. URL `http://jasss.soc.surrey.ac.uk/5/1/3.html`.

Grimm, V. (1999). Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological Modelling* **115**, 129–148.

Grimm, V. (2008). Individual-based models. In: *Ecological Models* (Jørgensen, S. & Fath, B., eds.), vol. 3 of *Encyclopedia of Ecology*. Elsevier.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J., Jørgensen, C., Mooij, W., Müller, B., Pe'er, G., Piou, C., Railsback, S., Robbins, A., Robbins, M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R., Vabø, R., Visser, U. & DeAngelis, D. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling* **198**, 115–126.

Grimm, V., Frank, K., Jeltsch, F., Brandl, R., Uchmański, J. & Wissel, C. (1996). Pattern-oriented modelling in population ecology. *The Science of the Total Environment* **183**, 151–166.

Grüne-Yanoff, T. (2009). The explanatory potential of artificial societies. *Synthese* **169**(3), 539–555. (originally presented at Models & Simulation 1, IHPST, Paris 1996).

Howick, S., Eden, C., Ackermann, F. & Williams, T. (2008). Building confidence in models for multiple audiences: the modelling cascade. *European Journal of Operational Research* **186**, 1068–1083.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. University of Chicago Press.

Ladley, D. & Bullock, S. (2007). Integrating segregated markets. *International Transactions on Systems Science and Applications* **3**(1), 11–18.

Langton, C. (1987). Artificial life. In: *Proceedings of the Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems (ALIFE '87)* (Langton, C., ed.). Addison-Wesley.

LeBaron, B. (2001). A builder's guide to agent-based financial markets. *Quantitative Finance* **1**(2), 254–261.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist* **54**(4), 421–431.

Lorenz, E. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**(2), 130–141.

Manzo, G. (2008). Book review of Gilbert, N. – Agent-Based Models. Journal of Artificial Societies & Social Simulation 11:2. URL `http://jasss.soc.surrey.ac.uk/11/2/reviews/manzo.html`.

Marmiroli, M., Tanimoto, M. & Tsukamoto, Y. (2007). Revenue risk analysis for asset management. In: *CIGRÉ Osaka Symposium: System Development and Asset Management under Restructuring*.

McArthur, S. D. J., Davidson, M., Catterson, M., Dimeas, L., Hatziargyriou, D., Ponci, F. & Funabashi, T. (2007). Multi-agent systems for power engineering applications – part I: Concepts, approaches, and technical challenges. *IEEE Transactions on Power Systems* **22**(4), 1743–1752.

McKelvey, B. (2002). Model-centered organization science epistemology. In: *The Blackwell Companion to Organizations* (Baum, J., ed.), chap. 33. Blackwell Business, pp. 752–780.

Moss, S. (2008). Alternative approaches to the empirical validation of agent-based models. *Journal of Artificial Societies & Social Simulation* **11**(1), 5. URL http://jasss.soc.surrey.ac.uk/11/1/5.html.

Noble, J. (1997). The scientific status of artificial life. Poster at the Fourth European Conference on Artificial Life (ECAL97).

Noble, J., Bullock, S. & Di Paolo, E. A. (2000). Artificial life: Discipline or method? report on a debate held at ECAL '99. *Artificial Life* **6**(2), 145–148.

Polhill, J., Parker, D., Brown, D. & Grimm, V. (2008). Using the ODD protocol for describing three agent-based social simulation models of land-use change. *Journal of Artificial Societies & Social Simulation* **11**(2), 3. URL http://jasss.soc.surrey.ac.uk/11/2/3.html.

Read, D. (1990). The utility of mathematical constructs in building archaeological theory. In: *Mathematics and Information Science in Archaeology: A Flexible Framework* (Voorrips, A., ed.), Studies in Modern Archaeology. Helos, pp. 29–60.

Resnick, M. (1997). *Turtles, Termites, and Traffic Jams*. MIT Press.

Richiardi, M., Leombruni, R., Saam, N. & Sonnessa, M. (2006). A common protocol for agent-based social simulation. *Journal of Artificial Societies & Social Simulation* **9**(1), 15. URL http://jasss.soc.surrey.ac.uk/9/1/15.html.

Rossiter, S., Noble, J. & Bell, K. (2010). Social simulations: improving interdisciplinary understanding of scientific positioning and validity. *Journal of Artificial Societies & Social Simulation* **13**(1), 10. URL http://jasss.soc.surrey.ac.uk/13/1/10.html.

Roughgarden, J., Bergman, A., Shafir, S. & Taylor, C. (1996). Adaptive computation in ecology and evolution: a guide for future research. In:

Adaptive individuals in evolving populations: models and algorithms (Belew, R. & Mitchell, M., eds.), vol. 26 of *SFI Studies in the Science of Complexity*, chap. 2. Addison-Wesley.

Sallans, B., Pfister, A., Karatzoglou, A. & Dorffner, G. (2003). Simulation and validation of an integrated markets model. *Journal of Artificial Societies & Social Simulation* **6**(4), 2. URL `http://jasss.soc.surrey.ac.uk/6/4/2.html`.

Schelling, T. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology* **1**, 143–186.

Schick, T. & Vaughn, L. (2007). *How to Think about Weird Things: Critical Thinking for a New Age*. McGraw-Hill, 5th ed.

Shackley, S., Young, P., Parkinson, S. & Wynne, B. (1998). Uncertainty, complexity and concepts of good science in climate change modelling: are GCMs the best tools? *Climatic Change* **38**, 159–205.

Stanislaw, H. (1986). Tests of computer simulation validity: What do they measure? *Simulation & Games* **17**, 173–192.

Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill.

Swedberg, R., Himmelstrand, U. & Brulin, G. (1987). The paradigm of economic sociology. *Theory & Society* **16**, 169–213.

Voorrips, A. (1987). Formal and statistical models in archaeology. In: *Quantitative Research in Archaeology: Progress and Prospects* (Aldenderfer, M., ed.), chap. 3. Sage, pp. 61–72.

Werker, C. & Brenner, T. (2004). Empirical calibration of simulation models. Papers on Economics & Evolution 0410, Max Planck Institute.

Wheeler, M., Bullock, S., Di Paolo, E. A., Noble, J., Bedau, M., Husbands, P., Kirby, S. & Seth, A. (2002). The view from elsewhere: Perspectives on ALife modelling. *Artificial Life* **8**(1), 87–100.

Windrum, P., Fagiolo, G. & Moneta, A. (2007). Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies & Social Simulation* **10**(2), 8. URL `http://jasss.soc.surrey.ac.uk/10/2/8.html`.

Wooldridge, M. (2002). *An introduction to multiagent systems*. John Wiley & Sons.