JASSS

[Scott Moss](#) (2008)

Alternative Approaches to the Empirical Validation of Agent-Based Models

Abstract

This paper draws on the metaphor of a spectrum of models ranging from the most theory-driven to the most evidence-driven. The issue of concern is the practice and criteria that will be appro- priate to validation of different models. In order to address this concern, two modelling approaches are investigated in some detailed — one from each end of our metaphorical spectrum. Windrum et al. (2007) (http://jasss.soc.surrey.ac.uk/10/2/8.html) claimed strong similarities between agent based social simulation and conventional social science — specifically econometric — approaches to empirical modelling and on that basis considered how econometric validation techniques might be used in empirical social simulations more broadly. An alternative, the approach of the French school of 'companion modelling' associated with Bousquet, Barreteau, Le Page and others, engages stakeholders in the modelling and validation process. The conventional approach is constrained by prior theory and the French school approach by evidence. In this sense they are at opposite ends of the theory-evidence spectrum. The problems for validation identified by Windrum et al. are shown to be irrelevant to companion modelling which readily incorporates complexity due to realistically descriptive specifications of individual behaviour and social interaction. The result combines the precision of formal approaches with the richness of narrative scenarios. Companion modelling is therefore found to be practicable and to achieve what is claimed for it and this alone is a key difference from conventional social science including agent based computational economics.

Keywords:
Social Simulation, Validation, Companion Modelling, Data Generating Mechanisms, Complexity

Introduction

1.1

Although model validation has been an ongoing issue in the social simulation literature, there has so far been no systematic consideration of whether different approaches to validation are appropriate to different approaches to modelling and whether some validation approaches, and their associated modelling approaches, are preferable to others. Several recent papers in this journal afford us the opportunity to begin a comparison and evaluation of alternative validation and associated modelling approaches. The formulation or application of general epistemological principles is *not* an objective here. Instead, the intention is to describe and compare alternative

validation *practices*.[1]

1.2

The point of departure for the design and implementation of any formal model will be characterised by some balance of theory and evidence. Computational economic models are extreme in this regard since they incorporate computational implementations of utility theory, the economic theory of production and distribution, general equilibrium theory and/or game theory.[2] At the opposing extreme is *companion modelling*: an approach developed by Bousquet et al. (1999) (*e.g.* (1999)), where models are developed around evidence with no explicit theoretical starting point.[3] Companion modelling was a source of influence on models developed at different scales by Downing et al. (2000) and Geller and Moss (2007a). A general feature of companion modelling is that the models are designed and refined in a process involving the participation of stakeholders and other domain experts.

1.3

Windrum, Fagiolo and Moneta (2007) set out an account of model validation procedures conditioned by their perspective as agent based *economic* modellers. They claimed to elucidate important issues for agent based social simulation modellers as well. However, as will be argued below, validation techniques appropriate to economic modelling are only applicable in more restrictive conditions than Windrum et al. claimed.

1.4

Windrum et al. bring to their discussion a presumption which is natural for economists and (econo– or socio–)physicists (e.g. Weisbuch et al., 2000; Deffuant, 2006) and shared by many, possibly the bulk, of the agent based social simulation community as well. In the terminology of econometrics used by Windrum et al., a real data generating process is presumed to exist and a social model should capture relevant elements of that process without any biases due to the impact of elements not captured by the model.[4]

1.5

However, the real data generating process or mechanism seems to be something of a mystical concept. Social scientists of a post modernist persuasion deny its existence (cf. Vattimo, 1988). Econometricians, including Windrum et al., assert its existence but note that it cannot be observed directly — perhaps because not enough social data is collected. Either way, the best we can do in an attempt to validate agent based models as representations of a real data generating process is to compare simulation outputs with such data as is available. In essence, the model as data generating process is seen as being independent of whatever social processes generate the data we observe.

1.6

An alternative use of agent based models is to represent social processes as perceived by participating stakeholders. When used in this way, stakeholders participate by providing their own views and perhaps independent evidence about the nature of the environment, the reasons why different actors behave as they do, the structure and the nature of social interaction within the relevant community and, if appropriate, across communities. This approach to the design, implementation and exploitation of models is unique to quite a small segment of the social simulation community. Development of such models at local scale are due to Barreteau, Bousquet and Attonaty (2001), Barreteau, Le Page and Patrick D'Aquino (2003a) and Becu, Perez, Walker, Barreteau and Le Page (2003), among others. The team that, as far as I know, originated this approach has called it "companion modelling" (Bousquet et al., 1999) and subsequently published a manifesto for their approach in JASSS (Barreteau et al., 2003b). Companion modelling effectively embeds model development in the social process of policy or strategy development. Companion modelling on a larger social scale (southern England (Downing et al., 2000) or Afghanistan (Geller, 2006)) demonstrated and confirmed with *subsequently* acquired fine grain time series data that realistically specified agents' behaviour and their social interactions systematically generate unpredictable episodes of volatility in domestic water consumption and conflict intensities as measured by numbers of combat deaths.

1.7

The French school of social scientists who developed companion modelling at local scale have

not investigated the volatility issue. The social scientists who have engaged in companion modelling at wider scales have produced only a few examples of unpredictable volatility but no examples from which such volatility was absent. Moreover, companion models in which social networks emerge from descriptively accurate representations of individual behaviour appear to generate small world social networks (Alam et al., 2007; Geller and Moss, 2007b).

1.8

The purpose of this paper is to compare the validation procedures described by Windrum, Fagiolo and Moneta (2007) for agent based computational economic models with the validation procedures developed within a participatory framework by the companion modellers. Companion modelling does not employ representations of behaviour based on utility functions or social environments cast as iterated prisoners' dilemmas or social interaction as round robin tournaments. These modelling artifacts depend on the economists' and physicists' conception of the rle of modelling — a conception that is inapposite to modelling for purposes of social policy or strategy analysis and formation. For these purposes, companion modelling is appropriate not least because it avoids the very problems and issues identified by Windrum et al. The main reason for developing the companion modelling approach, however, is that — unlike the approach of economists and econo-physicists — it is useful and achieves what is claimed for it.

1.9

In order to achieve my purpose here, I will first offer a detailed critique of the approach taken by Windrum et al. I will then describe how the companion modelling approach avoids the pitfalls of the economic approach.

Validating agent based models: the economist's view

2.1

Windrum et al. begin by setting out the common ground between agent based social simulation and economics: models are bottom up in that they are comprised by heterogeneous agents; the agents have cognitive capacities, and the agents are socially embedded. The literature they cite to support their description is concerned with economic theory but the same points are frequently made in the social simulation literature where it is taken for granted that agents are different and learn and evolve differently through the course of a simulation[5]; the cognitive properties of agents are frequently important design elements[6] and social interaction is essential[7] Consequently, the claim by Windrum et al. to have specified agent based models in a manner that should be useful to all agent based social and economic modellers seems well founded. The further implicit proposition that real and model data generating mechanisms are independent is also widely, though not universally, shared by economic and social simulation modellers.

2.2

Windrum et al. claim that

> [s]ome AB economists, engaged in qualitative modelling, are critical of the suggestion that meaningful empirical validation is possible. They suggest there are inherent difficulties in trying to develop an empirically-based social science that is akin to the natural sciences. Socio-economic systems, it is argued, are inherently open-ended, interdependent and subject to structural change. How can one then hope to effectively isolate a specific 'sphere of reality', specify all relations between phenomena within that sphere and the external environment, and build a model describing all important phenomena observed within the sphere (together with all essential influences of the external environment)? In the face of such difficulties, some AB modellers do not believe it is possible to represent the social context as vectors of quantitative variables with stable dimensions…. (Windrum et al., 2007, paragraph 4.2)

2.3

It will be useful to unpack this paragraph in order to avoid distorting positions regarding the various virtues and vices of qualitative and numerical modelling and also the position with regard to the natural sciences.

2.4

Starting at the end, it is always *possible* to represent social context or any other social phenomena "as vectors of quantitative variables with stable dimensions." The question is whether some elements of social context and other social and cognitive phenomena are not better — that is, more accurately and precisely — represented by variables in a linguistic domain. Arguably, the answer to this question turns on the purpose of the modelling.

2.5

Windrum et al. assert that "it is argued" that socio-economic systems are "inherently open-ended, interdependent and subject to structural change." I am certain that Windrum et al. would not take issue with the claim that societies are interdependent since that is one of the three key characteristics they identified for all agent based modelling. It would also be surprising if they were to deny that societies are subject to structural change since major changes have taken place within living memory in the former Soviet Union and in the technological environment affected by electronics and communications.[8] If "open-ended" is a synonym for non-ergodic, then the authors also accept that feature of societies. Presumably then the objection is to the rejection of the possibility "effectively to isolate a specific 'sphere of reality' " in order to produce a model at all.

2.6

Finally, if it is the case that the "open-ended, interdependent and subject to structural change" features of society constitute "inherent difficulties in trying to develop an empirically-based social science that is akin to the natural sciences", should we deny those features of society, deny that they imply those inherent difficulties or simply ignore either the features of society or the inherent difficulties? Since Windrum et al. complain that neoclassical economists have effectively denied those social features, they must either be denying that they imply the inherent difficulties or they have chosen to ignore them.

2.7

These issues will be addressed in the context of the validation options described and discussed by Windrum et al. (2007, paragraphs 4.4-4.25). There are three such options: indirect calibration, Werker-Brenner calibration and history-friendly validation.

2.8

The steps in indirect calibration are:

1. Identify macro level "stylised facts" such as firm-size distributions or employment-growth relations.
2. Inform model design by "empirical and experimental evidence about microeconomic behaviour and interactions."
3. Use "empirical evidence on stylised facts" to restrict the parameter space — presumably in Monte Carlo studies, though this is not stated explicitly.

2.9

The Werker-Brenner calibration steps are:

1. Use existing empirical knowledge to calibrate initial conditions and the ranges of model parameters.
2. Obtain simulation outputs for each set of parameter values for the model.
3. Discard all sets of parameter values except those "that are associated to the highest likelihood by the current known facts (i.e. empirical realisations)."
4. Use the surviving parameter sets *together with domain expertise from "historians"* to further constrain the parameter space of the model.

2.10

Finally, the history-friendly steps are:

1. Design the agents and interaction mechanisms on the basis of detailed empirical studies, anecdotal evidence and historical studies.
2. Use this data to assist "the identification of initial conditions and parameters on key

variables likely to generate the observed history."

3. Use the data "to empirically validate the model by comparing its output (the 'simulated trace history') with the 'actual' history of the industry."

From period to period, the quality of the design that a company is able to achieve in each relevant dimension—performance and cheapness—improves according to the following equation:[2]

$$\text{change } X_i = a_0(R_i)^{\text{a1}}(T_j)^{\text{a2}}(L_i - X_i)^{\text{a3}}e \qquad (1)$$

[2] We use log linear equations whenever plausible because most economists have a good feel for their behavior and their quirks. Also, while in most cases the behavior being specified obviously relates to particular firms, to avoid clutter we do not denote that specifically.

Figure 1. The strongly quantitative approach; reproduced from Malerba et al. ([1999](), p. 15)

2.11

The two calibration methods take it for granted that the models are numerical since the stylised facts are numerical and the discussion of the parameters is not consistent with the specification of unordered sets of linguistic tokens. And, according to Windrum et al., "the history–friendly approach is strongly quantitative" which, as indicated by figure [1](), it most certainly is – at least for economists. Moreover, as figure [1]() also shows, the "history–friendly" approach is also informed — perhaps dominated — by a criterion of what "most economists have a good feel for". So even this approach is in practice constrained by standard economic modelling techniques such as utility theory and a preference for log linearity and linear homogeneity of functions (see, for example, [Malerba et al., 1999](), p. 20).

2.12

It seems fair to conclude that all of the validation approaches described by Windrum et al. depend on theories or techniques that are selected independently of the evidence and prior to designing and implementing any specific model.

Agent based social simulation practice: some evidence

3.1

Agent based modellers in the social simulation community produce models for a variety of purposes. Some models are produced to explore software design issues, some to produce computational representations of linguistically developed social theories, some simply to explore formal ways of describing social concepts such as norm and reputation and some to explore specific social issues. Empirical validation issues have been connected only with the latter.

3.2

If we consider just the issue of JASSS in which the Windrum et al. ([2007]()) paper appeared, there were seven refereed articles: the methodological paper by Windrum et al., two papers on financial markets ([Hoffmann et al., 2007](); [Peffer and Llacay, 2007]()), one on a simulation model of an actual non–financial–market exchange process ([Lee and Lee, 2007]()), an exploration of a minority game intended to describe social processes characterised by congestion ([Chmura and Pitz, 2007]()), a model of a proposed but unrealised resource allocation procedure ([Westera, 2007]()) and, finally a report of an experimental online scientific collaboration system ([Polhill et al., 2007]()). Four out of the seven papers were intended to represent actual social processes.

3.3

[Hoffmann et al. (2007)]() developed their agent specification from a set of prior theories of behaviour but then constrained those specifications on the basis of survey data from Dutch investors. They then conducted comparisons against macro level data from the Dutch stock exchange. This corresponds to indirect calibration. [Peffer and Llacay (2007)]() relied on a similar theoretical basis but without the direct evidential constraint of [Hoffmann et al.]()[9] and, indeed, without any attempt at empirical validation. Their purpose was strictly to explore the implications of various theoretical stances.

3.4

The remaining paper with a direct empirical element (Lee and Lee, 2007) derived its agent specifications as follows:

> To deduce real assumptions about sellers, we interviewed 40 grocery stores in the Seoul metropolitan area that sell perishable goods such as vegetables, fish, fruit, and dairy products. The store managers (or owners) were asked questions regarding list prices, their experience negotiating the price of perishable goods with buyers, and the appropriate range of product freshness levels, among other things. Thirty buyers (Seoul housewives) were also selected, and were asked to describe the utility factors they considered when negotiating with sellers. Based on these interview results, we established assumptions about sellers and buyers which we incorporated into our …experiment… (Lee and Lee, 2007, paragraph 4.5)

3.5

Of the two papers that relied directly on empirical evidence, both used that evidence to constrain the agent specification in the manner suggested by Windrum et al.. This amounted to stakeholder involvement but not stakeholder participation in that there was no attempt to engage the stakeholders in either design questions or in subsequent model validation. Moreover, in neither case was the interaction mechanism intended to be realistic. All organised financial markets are collections of brokers who place orders on behalf of their customers and jobbers who are required to own stocks of the securities in which they trade and to meet at some price any bids and offers from brokers or other members of their exchange — mainly, institutional investors. The interaction mechanism in the Hoffmann et al. paper was described by them in the following quote:

> The SSE [the artificial stock exchange] operates in the following four steps: (1) every investor in the market receives a personal signal (information on the next period's expected price) and observes the current market price, (2) depending on the confidence of the investor, the personal signal is weighted to a greater or lesser extent with the signal that neighbouring agents have received, and based on this an order is forwarded to the stock market, (3) a new market price is calculated based on the crossing of orders in the SSE's order book, and (4) the agent's rules can be updated according to their results. (Hoffmann et al., 2007, paragraph 2.3)

3.6

No discussion is offered about the relationship between the artificial and the real institutional arrangements for setting prices and negotiating or setting prices and quantities. Lee and Lee (2007) also devise a stylised representation of the exchange process and also do not discuss the relationship between the artificial and the real processes.[10]

Note that even the papers that specified agents on the basis of interview and/or survey data, converted their qualitative data to numerically defined utlty functions and the like to drive the agent behaviour. This can present problems when the results of the model are sensitive to the particular form and parameterisation of the functions (Edmonds, 2006) though, in response to this criticism, Deffuant (2006) argued that, in the face of such sensitivity, parameters and functional forms should be chosen to give the required result. If the required result is in conformity with some macro level social statistics, then Deffuant's response is what Windrum et al. would have offered.

3.7

There is, of course, no reason to expect a single issue of JASSS to be representative of all issues of the journal or the distribution of interests and approaches in the agent based social simulation field. Indeed, several types of important empirical study were not represented in the last issue. One is exemplified by the modelling of the actual Marseilles wholesale fish market by Weisbuch, Kirman and Herreiner (2000) who took pains to model the observed institutional fabric of exchange. Another, of course, is companion modelling and, in particular, the use of agent based models in rle play games as reported, for example, in a special issue of JASSS (see Barreteau et al., 2003a). Nonetheless, if we consider just that one issue of JASSS, the nature of empirical validation described by Windrum et al. seems to capture the validatory exercises undertaken in the empirical papers published in that.

Complexity and the purpose of validation

4.1

Complexity science is a loose-knit collection of interests centred around episodic and unpredictable volatility and the properties of mathematical networks. The first will be called *process volatility* and the second *structural volatility*. If the two are linked in a social context, individual behaviour that gives rise to the volatility will also lead to the establishment of small world networks. This connection has not yet been demonstrated systematically.

4.2

Unpredictable, episodic volatility[11], as has been observed for both physical (Jensen, 1998) and social (Moss, 2002) contexts, arises systematically in models where entities such as agents engage in fairly routine behaviour but will change that behaviour in response to strong stimuli, where the entities interact with some but not all other entities and where entities such as agents influence but do not universally imitate some or all of those with which they interact. Mandelbrot (1997), Fama (1963) and Moss and Edmonds (2005) have pointed out that neither classical statistical nor econometric theory can be presumed to be applicable in these conditions because (a) the variance of any underlying stable population distribution may not be defined so that (b) the law of large numbers will not apply. Between volatile episodes, individuals will mainly be engaging in routine behaviour so that econometric analysis might (this has not been proved) yield reasonable forecasts. Environmental scientists have a notion of an *x*-year event so that a 50-year event would be one that occurs on average every 50 years. But this does not mean that there is a reliable two per cent probability of a 50-year event occuring this year. We might not see such an event for several centuries and then observe a rash of them. The two floods of the Rhine in the early 1990s and none since is an example. Time series of water levels in the Rhine are a good example of distributions of relative changes with heavy tails – just as we observe not only in financial markets but also in supermarket sales of tea, biscuits, shaving preparations, alcoholic beverages and probably any brand of any fast moving consumer good (Moss, 2002). In all of these cases, econometrically forecasting the volatile events in real time simply does not happen. As stated in section 1, the same statistical signature and the same unpredictability pertains also to domestic water consumption and conflict. We do not know how widespread this phenomenon is, but neither to we know of any exceptions where it has proved possible to acquire data at a sufficiently fine grain of time step.[12]

4.3

The effect of process volatility is to limit the value of empirical validation as represented by Windrum et al.. Even if we allow for social inertia, there is no econometric means of identifying a probability of the occurrence of a volatile episode. Moreover, volatile episodes sometimes lead to long-run changes in the way people behave – to what econometricians call structural changes. Some of these changes are enormous such as those that result from major political upheavals and some are more modest such as those that result from volatile episodes in financial markets. But there is certainly no reason to believe that econometric descriptions of social processes during periods of widespread behavioural inertia will survive volatile episodes.

4.4

More generally, the structure of any model whether econometric or agent based that is validated against evidence acquired between volatile episodes cannot be presumed to survive subsequent volatility. So what then is the purpose of validation? Indeed, is there a purpose to model validation? I will argue that the answer to the second question is an emphatic affirmative and explore the purpose of validation below.

4.5

The position we face is that models with socially embedded, cognitively plausible agents cannot be used reliably to forecast the consequences of corresponding social processes. The reason is that such models produce unpredictable episodes of volatility in macro level time series and, where we have appropriately fine grain evidence, a similarly unpredictable episodic volatility is found in actual time series. Such social simulation models are therefore better candidates as representations of the real social data generating process than are models that do not produce such episodic volatility. Evidently, such models cannot be used to forecast turning points in trade cycles or financial market prices or other characteristics of social volatility. Whilst it is

certainly possible to condition models on data covering such episodes, even models designed econometrically to capture episodes of volatility (based on [Engle, 1982](#); [Bollerslev, 1986](#)) have yet to provide a correct forecast to such an episode. Moreover, it has long been known that models that perform well on data sets available at the time of their publication typically perform less well or badly when applied to post-publication data ([Mayer, 1975](#)). Between volatile episodes, it is possible that there is enough social inertia due to individual metastability and lack of social stress that econometric techniques can be tolerably effective in forecasting models. However, it is not possible to know when the next volatile episode will become manifest. Because of their reliance on stable distributions with well defined first and second moments, neither econometric nor classical statistical theory are at present able to alter that position.

Companion modelling

5.1

The proposition that neither econometric nor agent based nor any other known type of model can forecast volatile episodes or the state of society after such episodes helps to define the useful purpose of models and what types of models are useful.

5.2

Windrum et al. distinguish between models that are subject to empirical validation and models that have particular formal properties but are not open to validation. In their words,

> Certainly there are those … who have taken the step of accepting they are constructing and analysing synthetic artificial worlds which may or may not have a link with the world we observe (Doran 1997). Those taking this position open themselves to the proposition that a model should be judged by the criteria that are used in mathematics: i.e. precision, importance, soundness and generality. This is hardly the case with [agent based] models! The majority of [agent based] modellers do not go down this particular path. ([Windrum et al., 2007](#), paagraph 4.3)

5.3

This paragraph is worth unpacking. First, we need to be clear about the meaning of "precision, importance, soundness and generality". The textbook properties that any mathematical logic (or system) might have are consistency, soundness, decidability and completeness. We will turn presently to precision, importance and generality. A logic is consistent if it is not possible to prove any theorem or model of the logic is both true and false. A logic is sound if no false statement can be proved to be true. A logic is decidable if, given any well formed formula, there is a general and systematic method which can tell you if its true or false. A logic is complete if all statements that are true with respect to the logic can be proved to be true.

5.4

Furthermore,

> [L]ogical proofs embody certain constructions which may be interpreted as programs. Under this interpretation, propositions become types. …[I]n different contexts this is in fact an isomorphism: in a certain fragment of logic, *every proof describes a program and every program describes a proof*. ([Pfenning, 2004](#), emphasis added)

5.5

Since every agent based simulation model is a computer program then, by virtue of the *Curry-Howard isomorphism* described in the above quote by Pfenning, every model is a theorem of the programming language in which it is implemented. There is a fragment of logic that is consistent, sound, decidable and complete of which the simulation model is a well formed formula or is comprised of well formed formulae. This statement is true equally of numerical models of the sort widely used by agent based social modellers in general and also of models defined on linquistic domains as implemented in logic and declarative programming languages such as Prolog, Soar, DESIRE or SDML. All computer models are precise in the sense of being characterised by definiteness or exactness of expression. Because they are all logical proofs (if

they do not crash), every semantic statement of a model has a corresponding syntactical statement that is unambiguous. The process of proof renders unambiguous the relationships among statements constituting the steps of a proof.

5.6

In mathematics, the importance of a theorem is judged by the number of theorems that it supports. I will presume that Windrum et al. have something similar in mind – perhaps the number of citations to published reports of a model. Alternatively, we could consider a context-dependent view of importance. We could, for example, judge the importance of a model in terms of whether it supports stakeholders in forming expectations, in communicating with one another and in reaching decisions about actions to take.

5.7

The issue of generality raises a number of questions about the purpose of agent based modelling. One source of importance would surely be generality in the sense that a model is more important the greater the number of contexts in which, directly or indirectly by informing the design of other models, it supports decision making. However, there is no inherent reason why a model should be deemed unimportant if it supports decision making in one or a few specific contexts such as water management in a small catchment in Thailand (Becu, Perez, Walker, Barreteau and Le Page, 2003) or the Senegal River Valley (Barreteau, Bousquet and Attonaty, 2001).

5.8

It is clear from the following quotation that Becu et al. held a much less general and more qualitative view of validation (though they called it authentication) than do many ABSS researchers:

> In fact, the term validation is no longer adequate, as many interactions are beyond such an experimental approach. Authentication seems a better approach, as it requires forensic abilities and witnessing.
>
> For example, we have crosschecked the simulated cropping pattern with the ones coming from remote sensing mapping. From 20 repetitions, the average proportions of the different crop types overlap with the actual ones with 80% accuracy. We have also compared the average simulated yields with those provided by local Thai Agencies. In the case of rice, soybean and onion, mean yields are simulated with, at least, 70% accuracy. From an economic viewpoint, the emergence of a small group of wealthy "entrepreneurial" farmers appears to correspond well with the actual situation emerging in the catchment. The continuous impoverishment of the Poor category is less realistic but it has already prompted further consideration, along with our Thai colleagues, about the role of credit in relatively poorly performing households.
>
> Concerning the Farmers profiles, part of the initial material was derived from field surveys. But it is crucial to have direct feedback from the stakeholders themselves regarding the social and individual rules implemented (Barreteau et al., 2001). This recognition by the concerned actors is the best-known authentication …. (Becu et al., 2003, p. 329)

5.9

The result is clearly important to stakeholders in northern Thailand and the model, implemented in CORMAS which has been used in many applications by Bousquet and colleagues, utilises techniques and experience developed in previous such exercises.

5.10

What we get from such modelling is a targeted model that lends precision to the accounts given by stakeholders and supports the integration of formalisations of qualitative accounts with biophysical models or even national accounting models. The precision is due in part to the process of computational modelling itself since the model is a program and, by Curry-Howard isomorphism, a logical proof which is inherently precise. It is also due to the fine grain nature of the evidence — in the words of Becu et al. the "direct feedback from the stakeholders

themselves regarding the social and individual rules implemented".

5.11

Precision was not actually an issue for Windrum et al.. They did not consider precision in their discussion of validation except to dismiss it as a consideration for "the majority of modellers". Their explicit concern was the relationship between society as a data generating mechanism and the model as a data generating mechanism. That is to say, their concern was not with precision but with accuracy. For [Becu et al.](#) or [Barreteau et al.,](#) however, the only accuracy that was of concern was the accuracy with which they represented the views of stakeholders. But different stakeholders might well have different views and understandings of their own behaviour, the behaviour of others and the ways in which stakeholders interact with one another. These are not matters of some objective truth. The very mystery of the real data generating process renders it difficult and maybe impossible to identify with the sort of precision that is essential for minimal accuracy.

5.12

Evidently, models lend precision but not some sort of objective or global accuracy to social analysis. Agent based models where the agent design conforms to detailed field generated evidence (from surveys, interviews and rle playing games) are validated for the accuracy of their representations of the views of specific stakeholders. If different stakeholders have different views and understandings and even concepts of the world, then any one model might be an accurate representation of some stakeholders' views but an inaccurate (though precise) representation of other stakeholders' views.

5.13

This is a problem only if the models are intended to provide forecasts (including conditional forecasts of the effects of particular policy measures). However, forecasting over periods long enough to include volatile episodes cannot be reliable and, as far as I know, has *never* been observed. This conclusion is at odds with the conventional economic view (adopted by many social simulation modellers) that a model that is well validated against existing data can be expected to provide good forecasts in the sense that they would be equally well validated against future data. There is no evidence to justify that presumption and increasing simulation evidence to reject it.

5.14

The formal purpose of validation therefore, can only be for purposes of calibration and not for forecasting. The purpose of the models themselves is to introduce precision into policy and strategy discussions. The validation exercise integrates the models into the discussions of longer term processes helping the engaged stakeholders to disambiguate the terminology they use and to clarify their specification of the social processes generating the outcomes — the data — they anticipate. The models are no more likely to be in any sense true than are narrative scenarios and they lack the richness of the narratives. By integrating the modelling process into the development of narrative scenarios, policy and strategy analysts obtain the benefits of formal precision *and* the benefits of the rich expressiveness of storylines and scenarios.

Core methodological issues

6.1

Windrum et al. specified six core methodological issues. In this section, we will evaluate the effect on these issues of embedding the modelling process in a wider social process of policy and strategy formation.

Concretisation *vs* isolation

Essentially, a model is more concrete as it captures in more detail the entities contributing to the social process and it is more isolated as it reduces the set of such entities in order to concentrate on particular causal mechanisms. Windrum et al. identify two "open questions" in this regard. These are:

- How can we assess that the mechanisms isolated by the model resemble the

mechanisms operating in the world?

- In order to isolate these mechanisms, can we make assumptions that are 'contrary to fact', i.e. assumptions that contradict the knowledge we have of the situation under discussion?

When models are embedded in a participatory planning process, these questions are of little relevance. The entities to be included and how they are to be included (whether as agents or as patterns of interactions amongst agents) are issues for discussion with the planners and perhaps other stakeholders. It is the views of these principals that determine whether "the mechanisms isolated by the model resemble the mechanisms operating in the world". It is also very unlikely that, in these circumstances, stakeholders would knowingly agree to "assumptions that contradict the knowledge we have of the situation under discussion".

As-if assumptions

The issue here is whether "phenomena [should] behave *as if* certain ideal conditions were met: conditions under which only those real forces that are isolated in a model are active" ([Mäki, 2005](#), p. 501).

This is not an issue for a modeller designing and validating a model with stakeholder participation. The stakeholders and modeller agree on the forces that are to be isolated and then assess the outputs from simulation experiments with the model. If the outputs are not deemed plausible or the numerical outputs carry statistical signatures that are inconsistent with real social statistics, then one consequence might be a reassessment of the "real forces" (*i.e.*, forces considered to be real by participating stakeholders). The identification of the important forces is a part of the discussion and is an important element in the construction of policy scenarios.

Strong *vs* weak apriorism

"Apriorism is a commitment to a set of *a priori* assumptions. A certain degree of commitment to a set of *a priori* assumptions is both normal and unavoidable in any scientific discipline. This is because theory is often developed prior to the collection of data, and data that is subsequently collected is interpreted using these theoretical presuppositions." ([Windrum et al., 2007](#), paragraph 2.3) The distinction between "strong" and "weak" apriorism is that the latter "allows more frequent interplay between theory and data."

Embedding the modelling process in a policy process with stakeholder participation *always* puts evidence before theory. There is no *commitment* to a priori assumptions. All assumptions, whether drawn from theory or from prior modelling experience, are provisional and subject to abandonment or modification in the face of stakeholder-provided or other evidence. Windrum et al. would probably consider this to be weak apriorism. However, it is not an option but is rather in the very nature of companion modelling.

Analytical tractability *vs* descriptive accuracy

This is really an issue for economists since "[t]he neoclassical [economic] paradigm comes down strongly on the side of analytical tractability, the [agent based] paradigm on the side of empirical realism" and we are all agent based modellers.

The identification / under-determination problem

The issue here is that "different models can be consistent with the data that is used for empirical validation." Econometricians call this the identification problem and philosophers of science call it the underdetermination problem. Econometricians address the problem by adding equations to their models until the rank of the coefficient matrix is equal to the number of independent variables. The equations are based on economic theory.

The proposition that other models might be validated equally well on the available data is

not an issue because embedded models are only intended to represent individuals, their behaviour and social interactions as perceived by the participating stakeholders. If different stakeholders have different perceptions, they believe their respective models correctly capture those perceptions and the models are equally well validated in relation to macro level quantitative data, then the differences are not a problem for the modeller but for the stakeholders. Perhaps in such circumstances, the models and simulation outputs will not help to resolve issues. Perhaps the differences in the models might still be a useful subject for discussion by the stakeholders. Perhaps in such cases, conflict is more apparent than real. Whatever may be the case, the problem is not one of modelling except to the extent that the modelling exercise might be of no practical use.

The Duhem-Quine thesis

states that any hypothesis will implicitly rest on other - auxiliary - hypotheses and it is not possible to test any one hypothesis in isolation. Windrum et al. suggest this is a problem for strong apriorism which, as indicated, is not an issue for agent based modelling embedded in a participatory stakeholder process.

6.2

Evidently, none of the six core issues for validation as seen by Windrum et al. are issues for embedded policy (or companion) modelling. They all stem from the notion that there is a an objective and real but unobservable data generating mechanism and that the purpose of any model is to represent elements of that mechanism in ways that generate some of the same data and no data that is inconsistent with outputs from the real mechanism. If, however, we treat the purpose of the models to be the representation of perceptions by policy analysts and other stakeholders in the relevant social processes, these issues lack practical importance.

🌍 Conclusion

7.1

Two opposing approaches to social modelling and validation were considered here. The first is a class of agent based models that has much in common with mainstream economic models. They incorporate utility functions; they employ numerical representations of phenomena and attributes naturally described in qualitative terms by the individuals being represented and by other stakeholders; they misrepresent social institutions such as markets and other forms of social interaction as centralised information exchanges or payoff matrices or round robin tournaments. The second is a class of models emerging from a process that is embedded in the social process of policy and strategy formation. Such models are typically couched in linguistic terms used by stakeholders rather than numerical variables convenient and meaningful only to modellers. The models are developed to facilitate stakeholder participation in the model design and validation process. They are intended precisely to represent the perceptions of stakeholders in order to bring clarity to scenarios built to explore the possibilities - the opportunities and dangers - of an uncertain future.

7.2

There are, of course, many other approaches to, and uses of, agent based modelling. At the evidence driven end of the spectrum, agent based modellers and archaeologists have developed models of past civilisations that were validated against sedimentary and other evidence from archaeological digs. The classic example is the study of the Anasazi tribe in present-day New Mexico by Axtell et al. (2002). At the theory- or technique-driven end of the spectrum is the work on opinion dynamics typified by, for example, Deffuant (2006). Whilst evidence-based models are naturally validated, I am not aware of any attempts at specific validations of opinion dynamics models. If we take seriously the issues raised by Windrum et al. and explored here, then an appropriate step for opinion dynamics modellers and others far from the evidence-driven end of the evidence-theory spectrum would be to identify appropriate principles for the validation of their models. As far as I am aware, this has not yet happened.

7.3

While theory-driven models as a group have been applied to a wide range of social issues and policy analyses, virtually all companion modelling has been undertaken in relation to environmental and agricultural domains. This is not strictly a matter of choice by the modellers

but a consequence of where the funding is available. If potential for absorption into the prevailing mainstream of social science is a criterion for awarding research grants, then it is hard to see any fundamental change in this situation. Doubtless, models are less challenging to mainstream social scientists if they are remote from concrete reality because they avoid natural linguistic domains for variables or because they actively adopt mainstream algorithms to generate decisions by agents. Rigorous assessment of the value of companion and other types of evidence–driven modelling as a key element in social analysis clearly requires their application to a much wider range of social phenomena and issues than has been possible so far.

Acknowledgements

Notes

1

One of the anonymous reviewers of an earlier version of this paper complained that the discussion did not start from general epistemological principles. Judging by contributions to online discussion lists (cf. the simsoc archive for the year 2000 at http://www.jiscmail.ac.uk/ and Conte et al. (2001)), this view is by no means unknown amongst writers with an interest in social simulation but, equally, it is not held universally. Since, I am concerned with the practice of social simulation rather than any abstract theory of validation, I build my argument on the specific practices advocated by Windrum et al. and followed by the companion modellers.

2

This list is not meant to be exhaustive but does not do serious violence to the field as represented by Judd and Tesfatsion's (2006) handbook on computational economics.

3

The methods and results are nicely reviewed by Barreteau et al. (2003a) and Barreteau et al. (2003b).

4

For a similar distinction and criticism by a social simulation modeller without economic training, see Edmonds(1999). The same distinction was made much earlier by Hesse (1963).

5

This is so much the norm and always has been in social simulation that it is hard to select classic examples. Epstein and Axtell (1996) and Conte and Castelfranchi (1995) are both less recent and more frequently cited than most in the field.

6

This goes back to Newell (1990) and the whole SOAR literature

7

A standard citation here is Granovetter (1985)

8

This is an issue that has been of considerable concern to a number of leading econometricians such as Clements and Hendry (1996).

9

The questions they explored, however, were well informed by others' empirical work, particularly MacKenzie (2003).

10

Though part of their purpose is to investigate a possible process that has not yet been realised.

11

For present purposes, we shall say that episodic volatility is unpredictable if no known econometric method has been used systematically and correctly to forecast volatile episodes occurring after production of the forecast.

12

The central limit theory implies that coarsening the grain of time step (monthly rather than daily data, for example) will tend to make the distribution of relative changes approximate the normal distribution. (See Mandelbrot, 1997)

References

ALAM S J, Meyer R and Edmonds B (2007) Signatures in networks generated from agent-based social simulation models. CPM Report 07–176, Centre for Policy Modelling, Manchester Metropolitan University Business School, Manchester.

AXTELL R L, Epstein J M, Dean J S, Gumerman G J, Swedlund A C, Harburger J, Chakravarty S, Hammond R, Parker J and Parker M (2002) Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl. 3), pp. 7275–7279.

BARRETEAU O, Bousquet F and Attonaty J M (2001) Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of Artificial Societies and Social Simulation*, 4(2), p. 5, URL <http://jasss.soc.surrey.ac.uk/4/2/5.html>.

BARRETEAU O, Le Page C and Patrick D'Aquino P (2003a) Role-playing games, models and negotiation processes. *Journal of Artificial Societies and Social Simulation*, 6(2), p. 10, URL http://jasss.soc.surrey.ac.uk/6/2/10.html.

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

BECU N, Perez P, Walker A, Barreteau O and Le Page C (2003) Agent based simulation of a small catchment water management in northern Thailand: Description of the CATCHSCAPE model. *Ecological Modelling*, 170(2–3), pp. 319–331.

BOLLERSLEV T (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, pp. 307–327.

BOUSQUET F, Barreteau O, Le Page C, Mullon C and Weber J (1999) An environmental modelling approach: the use of multi-agent simulations. In Blasco F (ed.) *Advances in environmental and ecological modelling*, Paris: Elsevier,, pp. 113–122.

CHMURA T and Pitz T (2007) An extended reinforcement algorithm for estimation of human behaviour in experimental congestion games. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 1, URL http://jasss.soc.surrey.ac.uk/10/2/1.html.

CLEMENTS M and Hendry D (1996) Intercept corrections and structural change. *Journal of Applied Econometrics*, 11(5), pp. 475–494.

CONTE R and Castelfranchi C (1995) *Cognitive and Social Action*. UCL Press.

CONTE R, Edmonds B, Moss S and Sawyer R K (2001) Sociology and social theory in agent based social simulation: A symposium. *Computational and Mathematical Organization Theory*, 7(3), p. 183.

DEFFUANT G (2006) Comparing extremism propagation patterns in continuous opinion models. *Journal of Artificial Societies and Social Simulation*, 9(3), p. 8, URL

http://jasss.soc.surrey.ac.uk/9/3/8.html.

DOWNING T E, Moss S and Pahl Wostl C (2000) Understanding climate policy using participatory agent based social simulation. In Moss S and Davidsson P (eds.) *Multi Agent Based Social Simulation*, Berlin: Springer Verlag, *Lecture Notes in Artificial Intgelligence*, volume 1979, pp. 198–213.

EDMONDS, B. (1999). *Syntactic Measures of Complexity*. Doctoral Thesis, University of Manchester, Manchester, UK. http://cfpm.org/~bruce/thesis

EDMONDS B (2006) Assessing the safety of (numerical) representation in social simulation. In Billari F, Fent T, Prskawetz A and Schefflarn J (eds.) *Agent-based computational modelling*, Heidelberg: Physica Verlag, pp. 195–214.

ENGLE R (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50, pp. 987–1007.

EPSTEIN J M and Axtell R (1996) *Growing artificial societies: social science from the bottom up*. Complex adaptive systems, Washington, D.C.; Cambridge, Mass.; London: Brookings Institution Press: MIT Press.

FAMA E F (1963) Mandelbrot and the stable paretian hypothesis. *Journal of Business*, 36(4), pp. 420–429.

GELLER A (2006) *Macht, Ressourcen und Gewalt: Zur Komplexit‰t zeitgenssischer Konflikte. Eine agenten-basierte Modellierung [Power, Resources, and Violence: On the Complexity of Contemporary Conflicts. An Agent-based Model]*. Zurich: vdf.

GELLER A and Moss S (2007a) The Afghan nexus: Anomie, neo-patrimonialism and the emergence of small-world networks. CPM Report 07-179, Centre for Policy Modelling, Manchester Metropolitan University Business School, URL http://cfpm.org/cpmrep179.html.

GELLER A and Moss S (2007b) Growing *qawms*: A case-based declarative model of afghan power structures. CPM Report 07-180, Centre for Policy Modelling, Manchester Metropolitan University Business School, URL http://cfpm.org/cpmrep180.html.

GRANOVETTER M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), pp. 481–510.

HESSE, MB. *Models and Analogies in Science*. London: Sheed and Ward, 1963.

HOFFMANN A O I, Jager W and Von Eije J H (2007) Social simulation of stock markets: Taking it to the next level. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 7, URL http://jasss.soc.surrey.ac.uk/10/2/7.html.

JENSEN H (1998) *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge: Cambridge University Press.

JUDD K L and Tesfatsion L (eds.) (2006) *Handbook of Computational Economics, Vol. 2: Agent-Based Computational Economics*. Handbooks in Economics Series, North-Holland.

LEE K C and Lee N (2007) Cards: Case-based reasoning decision support mechanism for multi-agent negotiation in mobile commerce. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 4, URL http://jasss.soc.surrey.ac.uk/10/2/4.html.

MACKENZIE D (2003) Long-Term Capital Management and the sociology of arbitrage. *Economy and Society*, 32(3), pp. 349–380.

MÄKI U (2005) Models are experiments, experiments are models. *Journal of Economic Methodology*, 12(2), pp. 303–315. Cited by (Windrum et al., 2007).

MALERBA F, Nelson R R, Orsenigo L and Winter S G (1999) 'history-friendly' models of industry evolution: the computer industry. *Industrial and Corporate Change*, 8(1), pp. 3–41.

MANDELBROT B (1997) *Fractales, Hasard et Finance*. Paris: Flammarion.

MAYER T (1975) Selecting economic hypotheses by goodness of fit. *: The Economic Journal*, 85(340), pp. 877–883.

MOSS S (2002) Policy analysis from first principles. *Proceedings of the US National Academy of Sciences*, 99(Suppl. 3), pp. 7267–7274.

MOSS S and Edmonds B (2005) Sociology and simulation: Statistical and qualitative cross-validation. *American Journal of Sociology*, 110(4), pp. 1095–1131.

NEWELL A (1990) *Unified Theories of Cognition*. Cambridge MA: Harvard University Press.

PEFFER G and Llacay B (2007) Higher-order simulations: Strategic investment under model-induced price patterns. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 6, URL http://jasss.soc.surrey.ac.uk/10/2/6.html.

PFENNING F (2004) *Lecture Notes on The Curry-Howard Isomorphism*. Carnegie Mellon University, Pittsburgh PA, URL http://www.cs.cmu.edu/~fp/courses/312/handouts/23-curryhoward.pdf.

POLHILL J G, Pignotti E, Gotts N M, Edwards P and Preece A (2007) A semantic grid service for experimentation with an agent-based model of land-use change. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 2, URL http://jasss.soc.surrey.ac.uk/10/2/2.html.

VATTIMO G (1988) *The End of Modernity: Nihilism and Hermeneutics in Postmodern Culture*. Baltimore MD,USA: Johns Hopkins University Press. Translated by Jon R. Snyder.

WEISBUCH G, Kirman A and Herreiner D (2000) Market organisation and trading relationships. *The Economic Journal*, 110(463), pp. 411–436.

WESTERA W (2007) Peer-allocated instant response (pair): Computational allocation of peer tutors in learning communities. *Journal of Artificial Societies and Social Simulation*, 10(2), p. 5, URL http://jasss.soc.surrey.ac.uk/10/2/5.html.

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