The Use of Models
- making MABS more informative

Bruce Edmonds

Centre for Policy Modelling,
Manchester Metropolitan University,
Aytoun Building, Autoun Street, Manchester, M1 3GH, UK.
b.edmonds@mmu.ac.uk http://www.cpm.mmu.ac.uk/~bruce

The use of MABS (Multi-Agent Based Simulations) is analysed as the modelling of distributed (usually social) systems using MAS (Multi-Agent Systems) as the model structure. It is argued that rarely is direct modelling of target systems attempted but rather an abstraction of the target systems is modelled and insights gained about the abstraction then applied back to the target systems. The MABS modelling process is divided into six steps: abstraction, design, inference, analysis, interpretation and application. Some types of MABS papers are characterised in terms of the steps they focus on and some criteria for good MABS formulated in terms of the soundness with which the steps are established. Finally some practical proposals that might improve the informativeness of the field are suggested.

Introduction

What is MABS? It could be:
- *Entertainment* – a sort of intellectual computer game where one sets up an artificial system with lots of agents and then play with it to see what sort of effects one can get;
- *Art* – multi-agent systems designed and/or constructed for others to admire and enjoy;
- *Illustration* – multi-agent systems designed to animate or otherwise illustrate some sociological, philosophical or mathematical principle, in other words, a sophisticated pedagogic tool;
- *Mathematics* – using simulation as a stand-in for symbolic deduction in distributed systems where such deduction is impractical;
- *Communication* – multi-agent systems as an interactive medium for social exploration, negotiation and communication; or
- *Science* – multi-agent systems as a tool for understanding observed systems.

All of the above are legitimate uses of multi-agent systems. Each have different goals. Each have different roles in society. Each have different criteria for success. It is not obvious that academics who attend MABS workshops have decided what MABS is. Indeed, it is not immediately obvious that there is a need to decide – these different activities can have much to contribute to each other.

However, it is likely that if these different activities are conflated together in a single paper then only confusion will result. These activities have very different aims, and so it is unlikely that they will be able to simultaneously satisfied at a high standard. Further if the aims of a piece of work are unclear there is a considerable danger that it will be misinterpreted – for example, if a MAS was designed as an illustration of some philosophical principle then it would be a mistake to take this as an indication of the behaviour of any particular natural system.

In *this* paper I will only examine the use of MAS as a tool for understanding – i.e. MABS as a science. In doing this I well understand that these other sorts of activities will be involved, but that the *ultimate goal* of the methodology I will be examining is to gain some understanding by people of systems that are observed. This does not mean that there will not be meaningful *sub-goals* (as will become clear) or that sometimes there will also be *other goals[1]* requisite upon any particular piece of academic work (e.g. when one needs a model to be developed *with* some stakeholders).

To do this I will introduce an abstraction of the scientific modelling process and use this as a *framework* for analysing the activity of MABS as a tool for understanding observed systems. I hope to show that this analysis is appropriate, does indeed characterise some of the papers in this volume, enables clearer judgments about their precise role and provides a basis upon which to judge their success. I conclude by applying this analysis in terms of practical suggestions that would make MABS more informative about observed systems, i.e. be a better science.

---

[1] I do not want to imply that these other goals are less important, in fact they are often *more* important.

An Analysis of Modelling

Several philosophers (for example, Hesse 1963, Rosen 1985 and Hughes 1997), who have observed the process of science have delineated the same picture of modelling. They see a model as enabling an inference process such that it corresponds to some aspect of an observed process. That is, if the inference process is set up so that its initial conditions are appropriate to observed system's initial state then the results of the inferential process will predict the relevant aspect of a later state of the observed system (to some degree of accuracy). This is illustrated in figure 1 below.
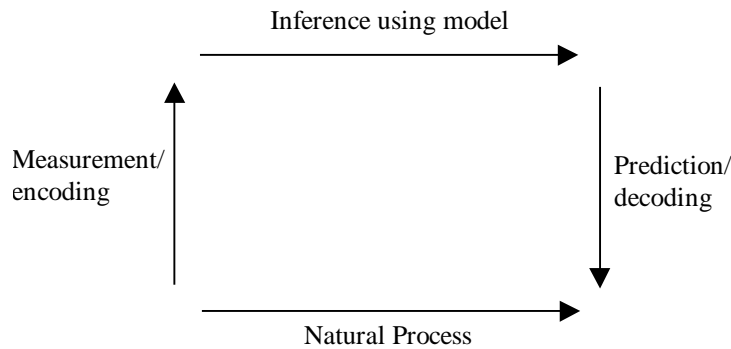
Inference using model

Measurement/
encoding

Prediction/
decoding

Natural Process

Fig. 1. The basic modelling relation

In this picture something *is* a model of something else if the above diagram *commutes*, that is to say that if you follow the two different routes around the rectangle then you get the same result.

Frequently in the natural sciences, the model has been encoded as a set of mathematical equations with numbers used as the basic currency for initial conditions and predictions. It was (and is) the job of mathematicians to discover and check the inferential machinery and the job of scientists to discover and check the mappings to and from the mathematical model.

In an imperative computational simulation the model is encoded as a program and the inference is performed by executing that program. In a declarative computational simulation the model is encoded as a set of logical statements and relations and the inference is done by an inference engine acting on these statements. In either case, it is the job of computer scientist to design and check the simulation process.

Understanding Multi-Actor Systems Through Modelling With MAS

Multi Agent Based Simulations attempt to model a multi actor system with a multi agent system. The modelled actors can be almost anything including: humans, institutions, agents, robots, programs, computers, objects, concepts, positions or even "balls" of water. Whatever the nature of the objects in the target system, the aim is the same: to understand the working of that system through the construction of MAS models and the analysis of their behaviour when run. The situation is illustrated in figure 2.
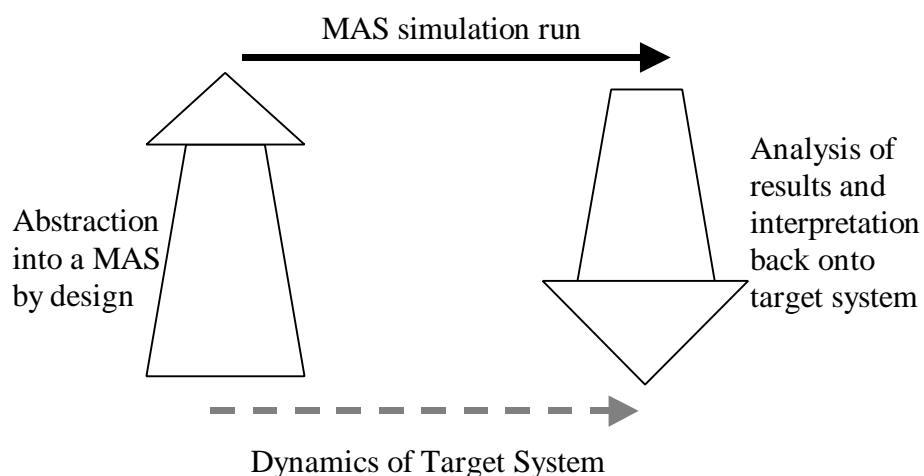
MAS simulation run

Abstraction
into a MAS
by design

Analysis of
results and
interpretation
back onto
target system

Dynamics of Target System

Fig. 2. A MAS used as a formal model

The basic sequence is this:

1. a target multi-agent process is chosen and an MAS is designed so as to incorporate the relevant aspects of the target system's structure (*design*);
2. the MAS is then run (this is basically a form of *inference*);
3. the resulting process is analysed by a variety of means, which can include: simple inspection, Monte Carlo, visualisation techniques, or statistics (*analysis*);
4. finally, the *point* of the whole exercise is to conclude something about the target system, by interpreting any conclusions about the behaviour displayed in the MAS runs back in terms of that system. This interpretation can be as strong as a concrete numeric prediction about the behaviour of the target system or as weak as an indication of possible behaviours expressed in qualitative terms (*interpretation*).

This basic sequence holds for almost any formal modelling and MABS is no exception to this. Of course, the above characterisation of the MABS modelling process is simplified, in part because only rarely is direct modelling attempted. Typically several other layers of models are involved, which are not always made explicit. For example, in many papers in this volume, it is not the target system that is modelled but rather an abstraction of that system which is related to the target system in a vaguer way. Thus we have the picture in figure 3. The MAS is a formal model of the abstraction and the abstraction is an analogical model of the target system (or a class of target systems).
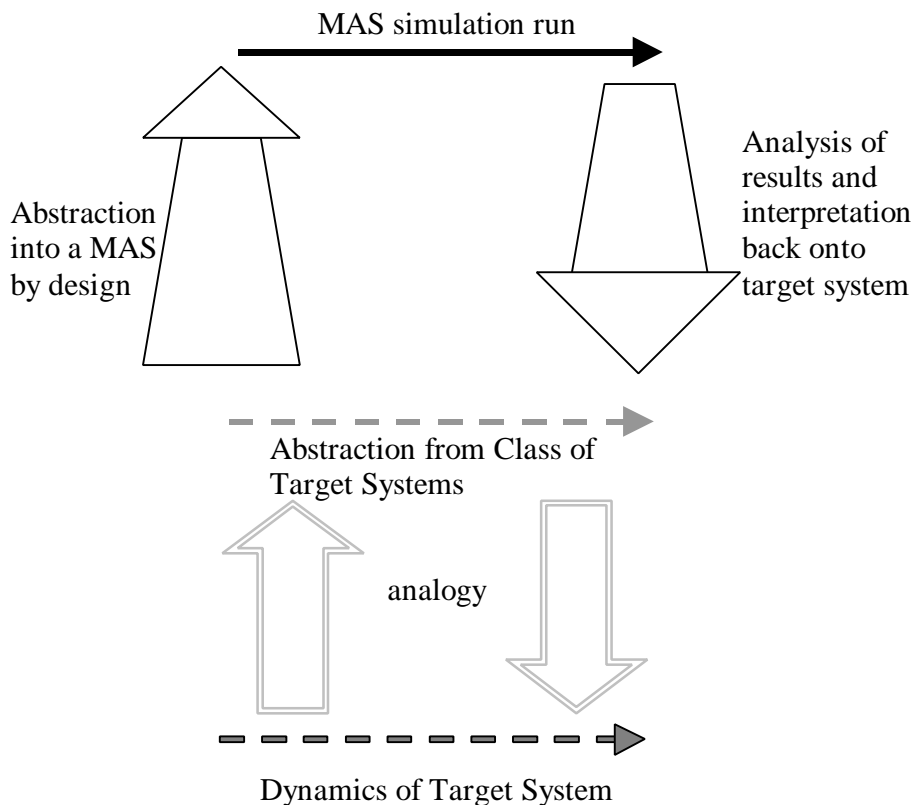

MAS simulation run

Abstraction into a MAS by design

Analysis of results and interpretation back onto target system

Abstraction from Class of Target Systems

analogy

Dynamics of Target System

Fig. 3. Modelling with an intermediate abstraction

In such a case there will be more stages to go through before we are able to infer something useful about the target system. First we abstract from the target system (*abstraction*); then we use MAS simulations to improve our understanding of our abstraction (*design*, *inference, analysis* and *interpretation*); and finally we apply our knowledge of the abstract system to the target system (*application*). This sort of "modelling at one remove" can be a very effective aid to decision making: the MAS modelling is used to hone one's intuitions and this increased understanding is implicitly utilised when making decisions with regard to the target system. This is what (Moss, Artis and Omerod 1994) discovered occurred in the development and use of UK macroeconomic models – the direct predictions of the models were almost completely useless on their own, but could be helpful to the experts that used them as an aid for thinking about the issues. The most common error in this sort of modelling is that modellers conflate the abstraction and the target in their minds and attempt to interpret the results of their MAS model directly in terms of the target system. This can manifest itself in terms of overly strong conclusions in

MABS papers, for example in (Axelrod 1984) at the end of chapter 5 it says: *"In this chapter Darwin's emphasis on individual advancement has been formalized in terms of game theory. This formulation establishes conditions under which cooperation based on reciprocity can evolve …"*.

The determination of the relevant abstraction can itself be analysed into two parts: deciding on the language/theoretical framework in which the abstraction will be specified; and formulating the abstraction within that framework.

The human cost of searching amongst candidate formal frameworks/languages is sufficiently great that one person usually has effective access to only a limited number of such frameworks. For this reason the framework is rarely changed once it has been chosen, rather people will go to considerable lengths to reformulate within a framework rather than change it. Added to this, the framework affects one's perceptions of the problem itself so that people 'see' problems in terms of their favourite framework (Kuhn 1962). Both of these effects mean that instead of the domain (i.e. the target systems) determining the most appropriate framework, the formulation is adjusted so that the target systems can be mapped into the framework even if this means distorting the target system almost beyond recognition.

Attempting to Improve Insight

Given the basic situation as described above, the question arises as to how one can improve the insight into the target systems in such modelling both in terms of the relevance and reliability of any conclusions. One way of approaching such a task is to improve any of the individual steps involved: abstraction, design, inference, analysis, interpretation or application.

By Increasing Strength of Formal Modelling Steps

The 'hard' sciences have often been characterised as concentrating on strengthening the formal modelling steps, and perhaps the most developed of these is the inference step. Utilising formal modelling techniques (including computational, numerical and symbolic models) ensures the reliability and consistency of this step. In conjunction with the rest of the formal modelling steps (design, analysis and interpretation), this forms the core of the modern "scientific method".

The danger of this approach is that improving the reliability of the formal modelling may be at the expense of its relevance to the target systems. The strength of the formal steps may be increased but the overall chain has been weakened because this is more than offset by the increased irrelevance of the abstraction to the target systems. In other words, although part of the chain is strengthened the overall chain has been weakened.

A hallmark of many 'degenerate research programs' (Lakatos 1983) is that they concentrate almost entirely on the purely formal aspects at the expense of their relevance to the target systems in their domain. It seems to be that a lack of empirical or practical success motivates a retreat into pure formalism. Particularly unsuccessful fields may retreat so far that the field concentrates upon a single step: the inference step. Such fields are marked by glut of purely formal papers and total irrelevance to any real systems – they lack the generality of pure mathematics *and* lack the relevance of good science.

By Improving the Correspondence with the Target System

Another well-established approach is to improve the analogical relations between the target systems and the abstraction, in other words to focus upon the descriptive aspects of modelling. Typically such an approach is marked by the presence of much richer and more complex abstractions, which lack the 'purity' of abstractions found in the 'hard' sciences. This approach is archetypal of the 'softer' social sciences[2].

This approach has the opposite pitfalls: it may be that the inference steps are so vague so as to be completely unreliable. This unreliability can mean that the abstractions do not even have sufficient coherency to allow a consensus about the meaning of parts of the abstraction in terms of the target systems to be developed. This can result in a situation where the academics involved are all using such terms in subtly different ways and discussions give the impression of "ships passing in the night". Again, it is the strength and relevance of the complete chain that counts.

---

[2] Although not always, (Wooldridge, 2000a) is a call to eliminate the abstraction-target system gap in agent-based computer languages, since for the most part the connection between formal systems (such as BDI logics) and the computer architectures (e.g. Mars) they where supposed to underpin is suggestive rather than strict.

It is notable that the natural sciences spend a huge amount of time and effort in establishing the correspondence of their abstract and target systems. This is not often apparent in a causal inspection of the output of these sciences, but is covered by all those involved in applied science. Indeed the activities of developing new measurement techniques, experimentally testing theory and applying theory in technology accounts for the greater part of the effort in such sciences. Thus although it has been a common perception that the hard sciences are marked out by their use of formal models, a deeper characterisation might point to the relatively developed and precise (though rarely simple – Cartwright 1983) correspondences between their abstractions and their target systems. It is this correspondence which is the engine of their success.

There are two ways to achieve greater descriptive correspondence: to choose an appropriate but specific framework; or to use a very expressive framework which embodies few and relatively weak assumptions about the subject matter. The first requires effective access to a relatively large number of frameworks and the conditions of application of each framework to be known. The second requires great expertise in the powerful framework so that a useful model can be formulated in it – this is a problem because the more powerful the framework the more difficult it is to calculate or prove anything. Thus the adoption of a very expressive framework can have negative effects such as shifting the difficulty to the design stage where one has to get a simulation to run, or again distorting the target system so that a relatively *simple* model can be fixed upon[3].

MAS as a Modelling Paradigm

What distinguishes MABS from other modelling enterprises is its use of MAS for the formal models. This represents a small step towards descriptive realism because it implies a commitment to analyse the target system in terms of the natural system boundaries that exist there. In other words it is almost universal in MABS to map objects, actors or other natural entities in the target system onto agents in the MAS, so that the 'boundary' of the entities correspond to those of the agents and that the interactions between entities correspond to interactions between agents. This is in contrast to when some agents are represented via average properties or single 'representative agents' (e.g. …). As in any modelling enterprise, the choice of MAS as the modelling paradigm will have consequences in terms of the reliability and relevance of the trade-offs described above. It does not, of course, avoid these trade-offs, but does allow a new trade-off to be exploited and, used well, can push the boundaries of both reliability and relevance forward (e.g. …). In the subsections below I summarise some of the consequences of adopting MAS as a modelling framework.

MAS as a Step Towards Greater Descriptive Accuracy and Precision

The use of MAS allows the application of formal models (computational models) to social systems (in its widest sense of interacting distinct entities) without the loss of relevance which has accompanied some attempts at such formal modelling (e.g. …). The MAS can be constructed so that the agents in the model and their interactions *do* correspond to those observed in the target systems. Thus the processes of abstraction and application become easier and more transparent.

Of course, MAS can easily be used so they do not improve either the reliability or the relevance of more traditional modelling techniques. The agents may not correspond to anything in the target systems, as happens in many evolutionary models where there is only a vague population to population relation, (Chattoe 1998). Another possibility is that the 'agents' in the model are *so* abstract that the relation between them and real target systems is, at best, merely suggestive.

More Difficult to Infer General Results

One consequence of the move to increased descriptive accuracy that is implicit in the use of MAS to model systems is the consequent particularity of any model outcomes. This lack of generality comes about in, at least, two ways. *Firstly*, the ability to build MAS models so that there is considerable structural correspondence between it and the target systems brings to the fore the generality-relevance trade-off. If the abstraction closely corresponds to a small class of target systems, then the conclusions can only be safely applied to those systems, and if the abstraction is more distant, that is, it attempts to capture some common aspects of a larger class of targets then the relevance of the abstraction to those systems becomes harder to establish and any conclusions less reliable. *Secondly*, there are inevitably some processes in the target systems (e.g. internal to the target

---

[3] Even a sufficiently expressive formal system can bias ones modeling as inevitably some sorts of model in that system will be simpler than others (Edmonds, forthcoming)

entities) that are not explicitly included in the model (either they are unknown, deemed irrelevant or impractical to implement[4]). It is a common tactic to substitute an indeterministic element into the MAS model to stand instead of this process, usually in terms of a random number or choice. In such a case it is not expected that an individual MAS trajectory will correspond to that of the abstraction (and hence of the target systems) but that the *collection* of trajectories corresponds to the possible trajectories of the abstraction.

Greater Contingency in Inference

The fact that most MABS are not deterministic, for the reasons noted above, means that the inference represented by the running of the model is modulus the indeterminism in that simulation. In simple non-agent-based simulations it occasionally happens that the randomness cancels out when a suitably large number of runs are performed. In these cases a Monte Carlo approach can be used and the range and central tendency of simulation behaviour safely deduced. In almost any agent-based simulation (and almost all simulations I have actually come across) one can not safely assume that random elements introduced into the structure will be uniformly carried through into the outcomes, so that one can safely generalise about 'central tendencies', equilibria and the like[5]. This contingency means both that single runs of the simulation may be completely unrepresentative of the system's general behaviour (if indeed there is such a 'general behaviour'), and that performing many runs of a simulation and averaging the results may merely result in a misleading artefact. In such cases a painstaking examination of single runs is often necessary in order to distinguish what is happening in each so that one can begin to determine how to classify the simulation trajectories. The increased descriptive realism has meant that the simulation has imported more of the object system's behaviour, including its unpredictability and complexity.

Syntactic Complexity Can Imply a Different Language for Results as for Specification

The complexity of the interactions and internal processes of many MABS means that it is often impractical to trace the trains of causation backward to determine a small set of causes for observed behaviour. Rather, as one traces the computation back the formal causes multiply until it encompasses almost the whole computation and every agent in the system. This is an example of 'causal spread' (Wheeler and Clarke 1999) and is common in complex and distributed systems. Sometimes it is the case that a *better*[6] explanation can be made in terms other than that of the systems detailed computation (e.g. an increase in 'entropy'). This is a clear example of *emergent* phenomena: the phenomena are not easily explained from the specification and detailed computation of the system, but make sense only within a *new* framework for their representation. The practical import of this is that the analysis and interpretation stages of the modelling process require much more attention than in simple deterministic or stochastic mathematical models.

Greater Variety of Possible Models

The move to a more descriptive modelling stance that is implied by the use of multi-agent models, and the effective encapsulation of the agents means that there are many more models that are specified differently (for example with different agent learning or reasoning mechanisms) but have essentially the same results (from the point of view of the intended interpretation). In such a case there is no single 'correct' model, but a whole class of models that give adequate results. It has been traditional to choose the 'simplest' adequate model as the 'correct' one, but there is no reason to suppose that such a model would be more representative of the target system's behaviour when used in different domains. Such 'simplest' models have pragmatic value in terms of ease of use and didactic value but are not justified on the grounds of being a better (or more likely) indicator of the truth (Edmonds, in press).

   The practical upshot of this is that it is desirable to constrain the space of candidate models as much as possible. In particular there is a *greater* need for such constraints in MABS than in simpler types of models where there are fewer possible variations. This is why the processes of verification and validation are

---

[4] These three reasons are frequently conflated in reports of simulations and only rarely explicitly distinguished and documented.

[5] Interestingly this is true even of deterministic agent-based systems, because the complexity of the system means that many aspects of the simulation are practically unanalysable, except as effectively indeterministic elements. For example, this is true of a pseudo-random number generator – which is precisely its point.

[6] Of course, what constitutes 'better' is the core of a large philosophical dispute, which I do not have space (or inclination) to go into here.

particularly important to MABS: verification enforces the constraint of its intended *design* (i.e. its conformance to the abstraction of the target system); and the validation ensures that the resulting processes that result from running the models are acceptable in terms of outcomes.

Strengthening verification requires that the abstraction (on which the design of a simulation is based) is specified as unambiguously as possible, and that the actual implementation of the simulation can be checked against this specification. The weakest possible verification is when the abstraction is not described directly but is only implicitly indicated via descriptions of the target systems and the actual simulation. In such cases it is impossible to tell which features of a simulation are intentional and which are merely necessary in order to get it to run (Cooper et al. 1996).

Strengthening validation means checking the output of the simulation in as many ways as possible by comparison with the system goals or actual target system behaviour. The strongest possible validation involves checking hard output at all stages and levels of detail of the model against unknown data gained directly from target systems to within pre-declared limits. The weakest validation is where the simulation is merely claimed to exhibit some qualitatively described behaviour. Validation which merely checks that the simulation behaves as intended does not add any additional constraints than those already introduced by verification, because it is essentially a design check.

Some Archetypes of MABS papers

To illustrate the analysis above, I present some archetypes of the sort of papers one finds at MABS conferences and journals followed by an analysis of an example of each taken from this volume[7]. These are in the order of the modeling steps that they focus on. They are also, in my opinion, roughly in order of increasing usefulness and rarity. Unfortunately the former have the academic status because they are more often cited in reports of the latter even if they have, in fact, contributed little to their success. As a generalization the field spends far too little effort on the later stages of the modelling process (analysis, interpretation and application).

Establishing an Abstraction

This sort of paper seeks to establish that a certain way of framing/describing/formalizing aspects of the target systems is necessary. The abstraction can be justified on the basis of either: *a priori* reasoning (their own or others) or on case studies of the target systems. Good examples of this sort of paper give indications of *how* they might be modelled in a simulation and *when* the proposed abstraction is appropriate. This sort of paper is only finally justified when the abstraction it proposed leads to simulations that, in their turn are informative about the target systems. Common forms of this include: proposals for abstract architectures and formalisations (e.g. a logical formalisation of trust).

A clear example of this type of paper is (Sawyer, 2000). This suggests a framework for modelling collaborative emergence in social systems based upon the thought and observations in sociology over the last century. This is primarily an argument as to what is *required* in order to model collaborative emergence via downward causation, with suggestions as to how such a simulation model might be designed. There is no suggested validation test for the success of such a simulation, but rather is motivated by an example of such collaborative emergence of dialogue from an improvised theater-piece exhibited. A couple of applications for a simulation incorporating this structure are suggested so an implicit validation for this abstraction is whether this does turn out to be a useful way of implementing these applications. There is no attempt to define the range of systems that this will be useful for except that it pertains to 'natural social systems' – a category which is very broad. It is thus difficult to build upon the work in this paper: one does not have a clear idea of the assumptions that are necessary in order for the stated conclusions to hold. It would be helpful to give an indication of *when* it is necessary to include the sort of structures described in this paper in a system (unless the implication is that one *always* needs it for any complex social system); and it would also be helpful to be given some way of knowing if one has succeeded. One could attempt to build upon this paper by *guessing* that it might be useful, trying to use it to build a system for one of the suggested purposes and then analyse whether the suggested structure did, in fact, help.

---

[7] To be precise the examples are taken from the MABS2000 pre-proceedings, as I did not have access to the final accepted versions. I strongly suspect that the characterisations will still hold, I suggest that you check this by reading them.

Documenting a Design Proposal

Here the abstraction is more or less given (often in previous papers by the authors or others) and a simulation design is proposed along with its justification. The merit of this sort of work is where is opens a way to implementing some aspect that was previously thought impractical. Its justification comes when simulations built upon the design come through with the goods, in other words the simulations do capture the aspects claimed for it in a verifiable way, they run with reasonable use of computational and human resources, and give results that under analysis are informative about the behaviour of the abstraction.

An example of this type of paper is (McGeary and Decker, 2000). This basically documents the design of a simulation of a Food Court. The abstraction of the food court and the system/agent architectures has already been decided upon and this document brings these two elements together. This abstract design is illustrated by one example: a more specific design for the process by which a Virtual Food Court (VFC) hires a (virtual) waiter. This design is situated by giving a little of the background to their model. Its purpose is to provide a system which people could use to experiment with in order to "*...explain, step-by-step, some specific and particular economic phenomena ...* [because we believe it] *is necessary to detail the conditions under which the studied phenomena happen ...*" (section 3 second paragraph). The proof of this will be if and when their goal is achieved using this system. It is difficult to see how this paper helps other researchers – there is not enough detail in it to allow others to duplicate it, and it is unclear which of the structural assumptions made derive from observations of real food courts, waiters etc. which derive from other people's theories and which are merely necessary in order to make implementation feasible. In the end, all it seems to do is keep others informed of what they are thinking so that they stay interested until they actually produce systems or get some results.

Exploring the behaviour of an abstraction

Here some aspect of an abstraction's behaviour is explored using simulations. This is a sort of stand-in for formal inference in analytic models. The relationship with target systems is typically only at the suggestive level. The utility of this sort of paper is (as in pure mathematics) dependent upon the *generality* of the results and knowing their *conditions of application* – if one could later on recognize another abstraction as sufficiently similar so that one can use the general results (for example as a guide to a simulation's design) this is helpful. If the results are quite particular to certain set-ups or the necessary structure for the discovered outcomes is not clear then it is almost impossible to *use* the results of such a paper. Sometimes such papers are used to make claims about the behaviour of target systems in a vague and ultimately unjustified way.

An example of this type of paper is (Axtell, 2000). All though some of the models described were obviously motivated and described in terms of observed systems, the objective of *this* paper is to demonstrate some general properties of MAS simulations: that: the topology of interaction; the medium of interaction; and the agent activation regime can each substantially change the results. Axtell is, in effect, saying *be careful about your simulation design and implementation you may be inadvertently biasing your results.* Although these results, stated in this way, seem obvious pointing them out with concrete examples is useful because may researchers *have* assumed that they were not relevant factors in their models[8]. This is a very general result, and a salutary lesson to researchers who want to over-interpret their particular results. However, in any implementation one is forced to make *some* such pragmatic decisions and it would be great if some *conditions* under which one had to be careful about such matters could be discovered.

Suggesting Solutions to Real Problems

Here the modelling cycle is finished in order to suggest some solutions to some real problems. That is conclusions about the behaviour and processes of an abstraction are applied to some target system in order to suggest solutions to some problem concerning them. In order for this to be done with any reliability there has to be a close connection between the abstraction and the target systems. If this connection is based upon a prior theoretical basis then the connection will be only as reliable as the extent of the practical validation of that theory, if it is based upon a descriptive basis then the application will only be meaningful in terms of the same descriptive language.

An example of this type is (Hemlrijk 2000). Here a concrete problem is focused on: explaining why in many group-living primates, males allow females access to resources exactly when they are sexually attractive. Here, an abstraction of the relevant primate behaviour is described, a MAS model produced, explored and analysed.

---

[8] It is notable that such 'errors' were often made because of an over-enthusiastic readiness to *apply* results from the analysis of a set of simulations straight to the target systems without regard from the difficulties of detangling the assumptions made in the abstraction, design and implementation of their simulations.

The conclusion of this is that there is a possible new explanation of this behaviour, namely that results from male attraction to females and not as a strategy to gain access to females. The model also makes some other predictions in terms of the levels of aggression which seem to match observations of such social systems. This paper is grounded in observations of real systems and completes the loop by applying its results back to those systems in a credible and restrained way. In addition to this the process of model building and the assumptions it makes are fairly clear. This work can be obviously built upon by building alternative models with the same or different assumptions and experimentally or observationally testing the predictions it makes.

Methodological Papers

A final type of paper is the one that discussed the methodology of MABS itself. This is in a slightly different position to those archetypes above because instead of doing MABS it is talking about *how to do* MABS. In fact, since MABS is still a young area of study, it is common for most papers to make *some* comment upon the methodology. The only justification of such papers is if it helps other researchers think about what they are doing, and do it better. Any amount of theoretical pronouncements about methodology is worthless compared to a single practical suggestion that is actually helpful.

An example of this type is (Edmonds, 2000). Personally I think it is crap, but then you will have to judge this one for yourselves.

Towards Criteria for Informative MABS

Given the characteristics of MABS as discussed above, the question arises as to how are we to distinguish the projects that will usefully inform us about the target systems that concern us. In other words, what are the criteria for good MABS work? We are not involved in *engineering* MAS to meet defined goals, so such considerations as the computational complexity of such design (as discussed in Wooldridge, 2000) are not directly relevant.

The ultimate criterion for any modelling enterprise is whether it *works* – in other words: does it help us to do things, predict outcomes in target systems, build systems that perform as we want etc. This is the *only final* justification of scientific intellectual work – even pure mathematics is justified by the fact that a lot of it *has* turned to be useful to other disciplines. Of course, it is notoriously difficult to predict which work will turn out to be useful, because many pieces of work are only useful when combined with other work. For example, one paper might specify a certain type of model and determine its behaviour and another might establish an application of this model to a target system (the traditional *theoretical* vs. *applied* distinction). The construction of knowledge about complex systems is necessarily a piecemeal and socially distributed project.

What we can do is to try to ensure that the complete modelling cycle is as strong as possible. This involves at least two areas. *Firstly*, that any modelling steps that are done (abstraction, design, inference, analysis, interpretation, application) are as sound as possible and as well documented as possible. This maximises the chance that one piece of work can be used by another, because the scope of its applicability will be identifiable and within this scope it is likely to be reliable. *Secondly*, that all the individual projects *do in fact* join up into a complete model of some target systems, i.e. that *all* the modelling steps are completed, are sound and connect up to form a complete chain.

Thus we can devise six process criteria that can help us judge work, one for each of the modelling steps.

1. **Abstraction**. *Is the abstraction specified?* Has it been made clear which aspects of the target system the abstraction is supposed to represent and over what class of target systems is it intended that the abstraction will cover? Is it clear that the abstraction corresponds sufficiently to the target system that it remains relevant?

2. **Design**. *Is it clear how the design relates to the abstraction?* Is it clear which parts of the simulation specification derive from the abstraction and which are details necessitated by the implementation? Is it possible to verify that the design does correspond to the abstraction and that the implementation does meet the requirements of the design? How do we know that there are no critical bugs in the simulation? How do we know that the important outcomes are not critically dependent upon the implementation language?

3. **Inference**. *Is the inference of outcomes sound?* To what extent are the outcomes described a necessary result of the model specification and design? Is the underlying inference process clearly specified and understood? Are the described outcomes representative of all the outcomes and, if not, what other types of outcome are there? Are the outcomes critically dependent upon particular parameter settings? Is it clear where indeterministic elements have been built into the model?

4. **Analysis**. *Is the analysis clear?* Has the analysis been well motivated in terms of what it abstracts from the outcomes and what it leaves out? Is the analysis replicable from the description? Is the raw data accessible

anywhere? Is the analysis technique demonstrably applicable to the outcomes? How reliable is the technique?

5. **Interpretation**. *Is the interpretation justified and relevant?* Do the results of the analysis justify the conclusions? Is the interpretation done into the same framework as the abstraction's specification?

6. **Application**. *Are the conclusions in terms of the target systems justified?* Do the strength of the other modelling steps provide the justification for the conclusions? Does the strength of the correspondence between abstraction and target system justify the strength of the conclusion?

It would be impractical for all these steps to be covered in every paper. Some papers will concentrate on one thing, e.g. the establishment of a suitable abstraction, and some will concentrate on other aspects, e.g. simulation analysis, etc. Of course, there is an onus to document and justify the steps that *are* covered. If a paper only focuses on some part of the complete modelling enterprise it also has an onus to specify two additional sections: its requirements in terms of the preceding steps that might lead up to it and its scope in terms of any steps that might follow from it. This will provide guidance to those who may be working on the missing sections.

For example, if a paper is focused on establishing an abstraction of some aspect of a class of target systems, then those who hope to use that abstraction in order to build a simulation and then to conclude something about outcomes of such an abstraction, will need to know what in the abstraction is essential and what a by-product of the descriptive process (this is especially important if the specification of the abstraction is formal). Those who hope to use known outcomes of a certain abstraction to draw conclusions about the target systems will need to know under what conditions they can safely do this (and indeed they will need to know what the target systems are).

Thus we can enumerate two more criteria.

7. **Conditions of Applicability**. *Are the conditions necessary for the described steps to be applied clear?* Are these conditions even known? Are the conditions directly specified or only implicitly indicated? Would a third person who read the paper know when they could use the work described?

8. **Generality of Conclusions**. *Is it clear in which circumstances the conclusions of the work hold?* Are all assumptions revealed and adequately documented? Is the reliability of the conclusions specified? Would a third person who read the paper know to when and to what the conclusions could be safely applied?

Conclusion: Some Ways Forward

With a view to improving the extent to which MABS can satisfy these criteria, I now conclude by making a number of practical suggestions. These are in addition to the obvious one of establishing a norm that published MABS papers meet the above criteria.

Strengthening Design Methodology

Despite the fact that many MABS papers concern themselves with the abstraction and design stages, the methodology in these areas is weak. This is contrast to the relatively large amount of attention that these steps have received in the wider multi-agent community. Here the ambition is that the abstraction should be specifiable in a logic with known properties and that an implementation that is based on this abstraction should be formally verifiable to that abstraction. So far this has been a more of a hope than a reality (with a few notable exceptions applied to 'toy' problems). It is unlikely that an abstraction that corresponded sufficiently to most of the target systems of interest to the MABS community would be amenable to such a complete verification, but that does not mean that there are not sensible steps that can be taken to strengthen these processes in our domain.

Firstly, as must be obvious from the comments above, there is a great need to explicitly distinguish and specify: what are the target systems of interest; what is the abstraction we are taking from these (including the framework and any assumptions or relevance judgments this involves); what is our intended design for our simulation; and finally what our implementation of our design is. These will all be, almost certainly, *different* – conflating them will obscure the conditions of applicability of the work described and thus impede their utility to others who might otherwise wish to build upon them.

This sequence of distinctions perhaps points towards a deliberately staged process from descriptions and perceptions of target systems up to an implemented simulation. Given the difficulty of verifying the sequence *in one jump*, the obvious thing is to separate it into sensibly small steps, each of which can be more carefully documented and understood. Some of these steps may be amenable to more formal verification and some will be composed of descriptive argument, but at least the *status* and *nature* of each step will be a lot clearer.

There are two tools that could assist in the atomisation of the abstraction-design-implementation process. *Firstly*, a "Social Simulation Specification Language" could help make the design of a MABS clear. This was suggested at a meeting of the Agent-Based Social Simulation (ABSS) SIG of Agent-Link but little progress has

been made towards its realisation. *Secondly*, one could envisage the extension of a constraint-based architecture which allows the early execution of fairly 'bare' specifications of simulations, albeit extremely inefficiently. This 'bare' framework could then be incrementally supplemented by additional constraints to make it gradually more efficient until we get to a fairly standard MABS simulation. The 'bare' framework will be a lot closer to the simulation specification and the incremental process could be checked so that it was apparent that the behaviour of the simulation at each stage was closely related to the previous one. It would also have the benefit of forcing out into the open any necessary 'assumptions' that were *not* part of the design. Some of the benefits of this sort of approach can be seen in some of Jim Doran's work (e.g. Doran 1997) and Oswaldo Téran's (Téran, Edmonds and Wallis, 2000).

Constraining Model Possibilities

As noted above, one of the effects of using the MABS paradigm is the explosion of possible simulation forms, with the implication that there will be many more simulations that display the same *outward* behaviour in terms of actions of agents. The upshot of this is that we need *as many* constraints upon our MAS models as possible and what constraints we apply need to be documented and checkable.

The most basic example of this is the *separate* checking of the simulation outcomes against a data model of the target processes, in addition to checks about the simulation structure and behaviour with respect to the design.

We should seek to verify and validate our models at as many levels of detail as possible. An example of this is proposed in (Downing, Moss and Pahl-Worstl, 2000) in which it is proposed that different levels of a social simulation be separately validated against the relevant level of the target systems and verified against each other. If a simulation is of an entity about which there is some knowledge about its internal computational processes, then these should be applied to the simulation unless a good reason can be presented. Maybe, at some point in the future we will know if and when a simpler, more efficient algorithm can be substituted for a complex cognitive process, but this is not the case at the moment.

Another way of increasing the constraints upon our models is by validating at a finer temporal level. That is to say that the intermediate stages of the resulting processes in the simulation should be checked against those of the abstraction or, even better, against data from the target systems. Just checking some statistics about the final outcomes is far weaker.

Strengthening the Generality of the Inference Step in MAS

The use of indeterministic elements in MABS simulations is almost always unavoidable and the systems are complex in fundamental ways. These two facts mean that simulations are highly contingent so that it can not be assumed that individual trajectories are *at all* representative. In addition to this the fact that simulations (as the systems they seek to model) can be highly *divergent* in their behaviour means that statistically established central tendencies and measures of spread can be very misleading. Thus we need some tools and methodologies for dealing with the whole *envelope* of simulation trajectories.

Firstly, the indeterministic elements in a simulation should be clearly tagged to indicate what sort of process they are 'standing-in' for (e.g. an unknown process, an irrelevant process, a computationally expensive process etc.).

Secondly tools are needed to control the indeterminism in a simulation and allow the systematic exploration of the complete *envelope* of the trajectories. This could involve a framework which allows the systematic exploration of possibilities in some sort of constraint-based model search (as suggested in Téran et. al).

Thirdly there could be established mappings between different architectures (MAS simulation, constraint-based search, theorem provers, statistical packages, data visualisation tools etc.) so that simulations could be transformed between them to enable different types of exploration to take place with respect to the same simulation design.

Structured Archive of MABS/SS Papers

One of the problems with the field is that there is very little work that builds on, compares or repeats other work. Notable exceptions include (Axtell, Epstein and Cohen, 1996) who attempted to 'align' models in different architectures and the "canonical task environments" of (Moss, 2000). What could help would be a structured archive of MABS results. If there could be a site where different types of MABS paper were deposited with enough information as to what they referred to being indexed it would be easier to start establishing and mapping out the connections between them. For example as each paper is submitted some extra information could be

entered to specify which: problems/target systems; abstractions/formal systems; modelling approaches/techniques; simulation implementations; sets of results; sets of analyses of results; and interpretations it refereed to. Each in each of these categories could be give a unique identifier, with new ones being added by authors as they felt necessary. Thus as well as some records enabling browsers to find connections between papers via their linking to these identifiers, a history and comments could be built up about each identifier representing each problem etc.

More Descriptive Modelling

Finally, I will end with a call for more low-level, descriptive modelling rather than ambitious abstract, high-level modelling that hopes to explain a lot *before* sufficient field work has been done. Historically 'armchair' theorising in advance of sufficient field work has not been successful[9], since our own preconceptions of how things should be is very strong. Rather we might be more successful if we attempt to produce generalisations in a *post hoc* manner, so that data and descriptions from our target systems can guide our modelling.

References

Axelrod, R. (1984) *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R., Epstein, J. M. and Cohen, M. D. (1996) Aligning Simulation Models: A Case Study and Results. *Computational and Mathematical Organization Theory*, 1:.

Axtell, R (2000) Effects of Interaction Topology and Activatioon Regime in Several Multi Agent Models, this volume.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford : Clarendon Press, 1983

Chattoe, Edmund (1998) Just How (Un)realistic are Evolutionary Algorithms as Representations of Social Processes?, *Journal of Artificial Societies and Social Simulation*, 1(3), http://www.soc.surrey.ac.uk/JASSS/1/3/2.html.

Cooper R., Fox J.& Farringdon J.& Shallice T. (1996) A Systematic Methodology for Cognitive Modelling, *Artificial Intelligence*, 85:3-44

Edmonds, B. (2000, forthcoming). Towards a Descriptive Model of Agent Strategy Search, *Computational Economics*. Also at http://www.cpm.mmu.ac.uk/cpmrep54.html

Edmonds, B. (2000) The Use of Models, this volume.

Edmonds, B. (in press), Complexity and Scientific Modelling, *Foundations of Science*. Also at http://www.cpm.mmu.ac.uk/cpmrep23.html

Doran, J. E. (1997) Foreknowledge in Artificial Societies" In Simulating Social Phenomena, In R Conte, R Hegselmann, and P Tierna (eds.), *Lecture Notes in Economics and Mathematical Systems*, 456:457-469.

Downing, T. Moss, S. and Pahl-Worstl, C. (2000) Integrated Assessment: Prospects for Understanding Climate Policy Using Participatory Agent-Based Social Simulation, this volume.

Hemelrijk, C. K. (2000) Sexual Attraction and Inter-secual Dominance amoung Virtual Agents, this volume.

Hesse, M. B. (1963) *Models and Analogies in Science*. London: Sheed and Ward.

Hughes, R. G. (1997) Models and Representation. *Philosophy of Science*, 64(proc), S325-S336.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1983) *The methodology of scientific research programmes*. Cambridge ; NY: Cambridge University Press.

McGeary, F and Decker, K. (2000) Modelliing a Virtual Food Court using DECAF, this volume.

Moss, S (2000) Canonical Tasks, Environments, and Models for Social Simulation, *Computational and Mathematical Organization Theory*, 6:249-275.

Moss, S. and Edmonds, B. (1998) Modelling Economic Learning as Modelling, Cybernetics and Systems, 29, 5-37. Also at http://www.cpm.mmu.ac.uk/cpmrep03.html

Moss, S., Edmonds, B. and Wallis, S. (1997) Validation and Verification of Computational Models with Multiple Cognitive Agents. CPM Report 97-25, MMU, 1997. http://www.cpm.mmu.ac.uk/cpmrep25.html

Moss, S., Artis, M. and Ormerod, P., A Smart Macroeconomic Forecasting System, *The Journal of Forecasting* 13, (3) 299-312, 1994.

Rosen, R. (1985) *Anticipatory Systems*. New York: Pergamon.

Sawyer, R. K. (2000) Simulating Emergence and Downward Causation in Small Groups, this volume.

Simon, H.A. (1986). The failure of armchair economics [Interview]. *Challenge*, 29(5), 18-25.

Terán, O., Edmonds, B. and Wallis, S. Mapping the Envelope of Social Simulation Trajectories. This volume. Also at http://www.cpm.mmu.ac.uk/cpmrep72.html

Wheeler, M. and Clark, A. (1999). Genic Representation: Reconciling Content and Causal Complexity. *British journal for the Philosophy of Science* 50: 103-135.

Wooldridge, M. (2000a) The computational complexity of agent design problems. Proceedings of the 4th International Conference on MultiAgent Systems – ICMAS-2000, Los Alamitos, CA:IEEE Computer Society, 341-348.

---

[9] As Herbert Simon pointed out about economics (Simon 1986).

Wooldridge, M. (2000b) Computationally Grounded theories of Agency. Proceedings of the 4th International Conference on MultiAgent Systems – ICMAS-2000, Los Alamitos, CA:IEEE Computer Society, 13-20.