Levins and the Lure of Artificial Worlds

Seth Bullock
Institute for Complex Systems Simulation, and
School of Electronics and Computer Science,
University of Southampton, UK
sgb@ecs.soton.ac.uk

Abstract:

What is it about simulation models that has led some practitioners to treat them as potential sources of empirical data on the real-world systems being simulated; that is, to treat simulations as 'artificial worlds' within which to perform computational 'experiments'? Here we use the work of Richard Levins as a starting point in identifying the appeal of this model building strategy, and proceed to account for why this appeal is strongest for computational modellers. This analysis suggests a perspective on simulation modelling that makes room for 'artificial worlds' as legitimate science without having to accept that they should be treated as sources of empirical data.

I

For practitioners across a growing number of academic disciplines there is a strong sense that simulation models of complex real-world systems provide something that differs fundamentally from that which is offered by mathematical models of the same phenomena. The precise nature of this difference has been difficult to isolate and explain, but it has sometimes been cashed out in terms of the ability to use simulations to perform "experiments" (e.g., Peck 2004). The notion here is that empirical data derived from costly experiments in the real world might usefully be augmented with data harvested from the right kind of simulation models. We will reserve the term "artificial worlds" for such simulations.

In this paper, rather than tackle the epistemological problems inherent in this type of claim head on, we will approach them obliquely by asking: what is the root of the attraction of constructing and exploring artificial worlds? By combining insights drawn from the work of Richard Levins, Valentino Braitenberg, Andy Clark and others, we arrive at an answer that at least partially legitimises artificial worlds by allocating them a useful scientific role, without having to assign the status of empirical enquiry to their exploration.

II

Our starting point in this exercise is simulation modelling work within the field of artificial life, where models, arguments and artefacts (robots, synthetic life, etc.) are used to explore fundamental biological questions (Bedau et al. 2000). The field as a whole has struggled with the notion of a special role for simulation models. Some argue that exploring the behaviour of artificial living things within the right kind of simulations constitutes a kind of "digital naturalism", as though software entities represent an additional kingdom of living creatures (Ray 1994; Lenski 2004). Others claim that simulations of artificial living systems are models, but that in the right circumstances they nevertheless have the capacity to settle empirical questions (Bedau 1998). Still others have cast these simulations as just another kind of reasoning aid to be placed in the scientific modelling toolbox alongside equational models, physical replicas, games, thought experiments, etc., being distinct from the others only in terms of their immaturity and in so far as they involve an added layer of indirection and complicatedness brought about by the involvement of automatic machinery (Di Paolo, Noble & Bullock, 2000).

As a locus for exploring the attraction of artificial worlds, artificial life simulations offer a number of key advantages: first, artificiality is overtly invoked in the (name of the) research

programme (Silverman & Bullock 2004); second, the research community has explicitly debated the role of simulation and its epistemological status (Wheeler et al. 2002); and third, crucially, it is generally accepted that most artificial life models lack theoretical security in the following sense. At one end of the security spectrum are *secure* computational models, each underwritten by mature and consensually agreed upon theory (sometimes to the extent that we do not explicitly recognise that there is a relevant underlying theory there at all). At the other end are *insecure* models that do not benefit from such a mature theoretical underpinning, but rather, are exploratory attempts to generate or progress such theory. Artificial worlds tend to arise and thrive in these less secure environs.

At the very most secure end of the spectrum, computational models do not look like models at all. Rather they resemble calculations or enumerations. Consider a rug and a room. Can the former fit inside the latter? The question might be explored computationally, but we might not be inclined to call the computer program a model. The theory underpinning the model is so deeply understood and relied upon that we do not think of it as theory. Transformations of shapes through rotation and translation, rules of enclosure or overlap: the ways that these ideas work for medium sized objects are no longer up for debate, they are settled. Consequently, a computer program written to explore the question of whether a particular rug will fit on the floor of a particular room seems to be calculating the answer rather than modelling the situation.

By contrast, a computer program that could calculate, not whether the rug fitted the room geometrically, but whether it fitted aesthetically would be navigating in less secure waters. We are not yet in possession of a mature and consensually agreed upon theory of human aesthetics and as a consequence we would not expect a simulation model to deliver reliable answers or predictions concerning whether the owners of the room would like or dislike a particular rug. We would perhaps be more inclined to imagine that models in the field of aesthetics would be used to test the internal consistency of theoretical positions, or to help identify experiments that might elegantly discriminate between different theories, etc.

Much scientific modelling lies somewhere between these two extremes, of course. Numerical simulation of fluid mechanics, for example, is relatively secure as a consequence of our strong faith in the Navier-Stokes equations as a basis for describing fluid flow. Consequently, the field of scientific computation is often concerned with efficiently implementing these equations rather than with testing them. Relatedly, although simulation models of global climate have to deal with much more than simple fluid flows, they inherit some of the theoretical security of computational fluid dynamics models and, partly as a consequence, have been able to play a significant role in informing climate policy. By contrast, simulation models of neural systems, for instance, are somewhat less secure since the theoretical basis for

nervous activity remains more open to debate. Perhaps uniquely, artificial life *deliberately* courts insecurity by identifying as its central explanandum the notion of "life as it could be" (Langton 1988), a deeply controversial and counterfactual theoretical postulate. The field of complex systems simulation is perhaps more characteristic of insecure modelling in general, attempting as it does to build models of real-world systems that are challenging as a result of the non-linear interactions amongst their many components. Here, there is typically explicit recognition of the underdeveloped theoretical basis for such models, and often but not always a sense that results may be more exploratory, theoretical and heuristic rather than predictive. However, it is in this type of territory that artificial worlds appear to be most attractive.

It is crucial to note that more secure simulations are not necessarily "better" than less secure simulations. For example, less than fully secure models surely play an important part in generating and refining the theory that will make future models more theoretically secure.[1]

Before beginning our effort to understand why artificial worlds are attractive to some scientists, we will consider one example in order to clarify the term. Early in his 1997 book "Would-be Worlds: How Simulation is Changing the Frontiers of Science", John Casti, a prominent complexity scientist, introduces an example of simulation modelling that is intended to be a compelling pointer towards the nature of the book's general thesis. He asks the reader to recall a recent American football Superbowl game in which one team beat the other by some margin of points. He then asks whether this result was a fluke or not. This is an empirical question about an event in the past. If at the time, we had asked the two teams to play a series of games, we could have perhaps arrived at an answer to the question by considering the set of outcomes, but it is not possible for us to go back in time and do this real-world experiment. However, Casti points to the existence of an American football computer video game that allows the player to control a team playing against a simulated version of either of the two teams involved. It is also the case that this video game allows two simulated teams to be pitted against each other (the player simply watches the simulated match unfold). Casti suggests that we might replay the Superbowl match a number of times within the game, collect the results and use these to settle the question of whether the real outcome was a fluke or not.

While Casti's example may strike us as whimsical or far-fetched, he is not alone in seeking or imagining simulations with sufficient validity to settle empirical questions or provide

[1] Given the examples used above, it might appear that model security tends to diminish in general as we move from physics through to the life sciences and environmental sciences, and on to the social, economic and psychological sciences. However, there are relatively insecure spots within any live discipline, including modern physics, and whether or not this apparent trend holds or not is unimportant for the argument in this paper.

pseudo-empirical data. Mark Bedau (1998) argues that a suitably programmed computer simulation would be able to settle the empirical thought experiment proposed by Stephen J. Gould (1989): if we returned in time to a point immediately prior to the origin of life on earth and gave the primordial soup a stir, would the subsequent, slightly disturbed, evolution of life on earth result again in life-forms similar to the ones around us today, or would wholly other forms arise each time the "tape" was re-run?

More substantively, several areas of scientific research are now pursued within "virtual laboratories" within which the use of models is treated as pseudo-empirical activity, e.g., from cell biology (Kitano et al. 1997) to archaeology (Premo 2006). Consequently, while this paper will draw on artificial life as a main source of examples, the issues that are dealt with here apply in varying degrees to simulation efforts within a much wider range of modelling disciplines that deal with systems for which theory is still under development, e.g., climate science, finance, medicine, neuroscience, complex systems, etc.

III

In his seminal 1966 paper "The strategy of model building in population biology", Richard Levins presents a trade-off between the aims that a modeller might attempt to meet in a single modelling exercise (Levins 1966). No useful model, Levins claims, can maximise precision, generality and realism.

Orzack & Sober (1993) have argued convincingly that a strict three-way trade-off between precision, realism and generality cannot hold since it is surely possible to decrease all three properties by simplifying a model. In line with Odenbaugh (2003), we would argue that Levins was not intending to introduce a strict trade-off of the kind that Orzack and Sober take issue with (and Matthewson & Weisberg 2009 analyse further). Moreover, here we argue that the dimensionality of Levins' argument has been misunderstood. Levins, in fact, explicitly invokes *four* interdependent model properties: "It is of course desirable to work with *manageable* models which maximize generality, realism and precision . . . But this cannot be done" (Levins 1966, p. 422, my emphasis).

Levins is arguing that increasing the realism, precision and generality of a model would not be sensible if this were achieved at the expense of what we will term a model's *tractability*: our ability to build and make use of the model productively.

[Figure 1 Here]

Levins introduces three modelling strategies that each avoid the threat of intractability by concentrating on only two of the other three remaining properties, labelling them only strategies I, II and III (see figure 1a). In doing so, he implicitly privileges tractability over the other three model properties since he describes how modelling strategies that sacrifice any one of generality, precision or realism may be useful, but omits a forth strategy that sacrifices tractability in an attempt to achieve all three of the other properties.

Barandiaran and Moreno (2006), apparently working independently of Levins' framework, introduce three useful labels that align well with the three Levinsian strategies, and a fourth label, which they associate with artificial life models and which occupies the middle ground that Levins avoids (see figure 1b). Strategy I seeks *mechanistic* models that strive for realism and precision. An example might be a model of the aerodynamic lift created by the splayed wing-tips of gliding birds (Tucker 1993). Strategy II seeks *generic* models that are not constrained by particular real-world systems but are general and precise in the sense that they are quantitative. An example might be the Watts-Strogatz small-world network model (Watts & Strogatz 1998) which generates networks that combine a significant degree of clustering with small characteristic path lengths between pairs of network nodes. Strategy III seeks *conceptual* models that are realistic and general but do not deliver precise quantitative outputs, instead shedding light on qualitative relationships amongst model parameters, variables or behaviours. An example might be a model of the evolutionary advantage of herding behaviour (Hamilton 1971) which must respect key real-world factors that constrain flocking creatures from a number of different species, but need not deliver a numerical value associated with the evolutionary advantage of flocking, or make precise predictions regarding when or where flocking will arise.

For Barandiaran and Moreno, a fourth category of modelling activity aims to generate what they term *functional* models which will be characterised here as lying somewhere between Levins' previously described three styles of model (see figure 1b). According to the authors, these models are "necessary tools to study a complex integrated behaviour". They are not *generic* models because they target the "behaviour or functionality exhibited by some natural system" and "must include constraints which are specific of the phenomena under investigation". They are not *mechanistic* models since they operate "where the [relevant] mechanisms ... are unknown, controversial or incompletely understood". They are not *conceptual* models since they are mechanistic to the extent that they are able to be used "to discover candidate mechanisms or local rules that produce or contribute to the observed and simulated global pattern or behaviour". As an example of this modelling style, Barandiaran and Moreno suggest models of insect swarm intelligence (e.g., Deneubourg 1991). An alternative might be models of the evolution of sexual reproduction (e.g., Watson 2006). Here, models must respect real-world constraints on evolutionary processes, and be quantitative (e.g., to the extent that they can address the "two-fold

cost of sex"), yet cannot afford to focus on the mechanisms of a particular species if they are to address the general form of the question: what enables or drives the evolution from asexual to sexual reproduction.

This fourth style of modelling described by Barandiaran and Moreno (2006) is clearly a hybrid of Levins' three strategies, often pursued in a space that lacks some theoretical security. As a consequence it is a strategy that flirts with intractability. That Levins felt a need to warn of the dangers inherent in attempting to build models that maximise precision, generality and realism is perhaps reason enough to assume that seeking to maximise all three properties might motivate some modellers. For our purposes, it is also instructive that although Levins presents the lure of these unmanageable models in rather general terms, he may have been thinking of simulation modellers as particularly appropriate recipients of his paper's message. Odenbaugh (2006) argues that, in writing his paper, Levins was responding to a growing trend in what he termed "Fortran ecology" (the use of large computer models that could be parameterised to represent a wide range of ecosystems and purported to generate precise, realistic predictions), which suggests that Levins may have considered that simulation modellers were particularly susceptible to the lure that he was warning against. But on the face of it, the attraction of maximally general, precise, realistic models should be felt quite generally across scientific modelling paradigms. Why would simulation modellers be particularly vulnerable to it?

IV

While the brute power of modern computational machinery is at the heart of the answer, Braitenberg's (1984) law of "downhill design and uphill analysis" offers us a more subtle account by halving the notion of tractability into components related to constructing and understanding, respectively. In his book, *Vehicles*, Braitenberg introduces some simple imaginary mechanical vehicles, the behaviour of which is governed by the way that their various sensors are wired up to their various motors. He describes how complex, apparently goal-oriented and intentional behaviour on the part of these vehicles can emerge from surprisingly simple internal designs. In doing so he points out that tinkering with the systems in order to achieve some interesting behaviour is far easier than analysing these systems to determine why a particular configuration of sensors, wires and motors gives rise to the particular behaviour that it does. Synthesis is more tractable than analysis.

For manually constructed mathematical models these two tractability halves tend to be tightly coupled, proceeding in step, and resulting in either an intelligible, working model or a

conceptually intractable dead end (i.e., a failed model). However, in simulation modelling this coupling is loosened until, for some simulations, construction *forgoes* understanding, meaning that simulations that "work" (in the sense of generating interesting or useful or empirically validated behaviour) can be arrived at well before they are fully understood.

This account is redolent of work by Clark (1990) on explanation in the context of artificial neural networks, where the automatic nature of the neural network algorithm can propel a modeller from a competence-level description of the problem to a working implementation of a solution without visiting the algorithmic level of representation necessary in order to achieve an understanding of that solution. By analogy, simulation modellers enjoy increased tractability in the design phase, relying on the automaticity of their models to produce interesting behaviour, but suffer during the analysis phase where model behaviour may remain analytically opaque unless significant effort is expended.

For example, compare the results of an imaginary equational model with those of a simulation model of the same phenomenon. The solution to a set of differential equations that constitute an equational model of some system might describe the conditions under which a particular class of equilibria exist. For example, the inequality $a < b$ might define a class of systems for which a certain strategy is stable. This solution is informative. It captures a relationship between two variables, a and b. Presumably, these variables have some understandable role to play within the model as a whole since they were initially invoked and defined by the modeller herself.

In contrast, a simulation designed to explore the same phenomenon as the equational model described above might result in several sets of data points that describe some aspects of the system as it varies over time, and over the space of possible parameterisations or variants of the model, and with the initial conditions from which the system evolves, etc. Upon sufficient analysis, these data points should reveal the same relationship as was discovered through the construction and analysis of the equational model, but the move from 'raw solution' (data points) to 'informative solution' (relationships) is more involved than the straightforward process described for the equational case outlined above.

Some effort must be made to reconstruct the *relationships between classes* that mathematical models utilise in explaining the behaviour of analytically derived models, from the *instances* that the simulation model generates.

Thus, although, under certain conditions, the construction of simulation models might prove more tractable than the construction of analogous equational models, the analysis of such

simulation models often requires an additional effort which threatens to more than compensate for any increased ease of design. This is Braitenberg's 'law of uphill analysis and downhill invention' for modellers, and has been given a more thorough exposition by Clark (1990) in terms of a distinction between automatic design processes and manual design processes.

Clark argues that manually designed systems are easy to interpret because the processes involved in their creation provide a natural way of decomposing them into intelligible sub-systems. For example, a machine manually designed to catch a thrown tennis ball might comprise various sub-systems each charged with effectively carrying out part of the ball-catching problem. Upon completion of the manually designed ball-catching system, an analysis of the manner in which it achieves its task is trivial (for the system's designer) since the manual design process involved specifying precisely this.

In contrast, a ball-catching system designed by an engineer using an automatic design process (e.g., harnessing some kind of artificial evolutionary process in order to design the tennis-ball catching system) is not so amenable to analysis. Although the engineer specified the problem, and the resources upon which the solution might draw (e.g., the general architecture of the system), she had no say in the manner in which the automatic process saw fit to exploit these resources, i.e., she had no part in constructing the algorithm that the automatic design process settled upon.

Clark (1990) cashes this idea out in terms of Marr's (1977) characterisation of design and explanation within cognitive science, incorporating Peacocke's (1986) addition of an intermediate "level 1.5" to Marr's original three-tier "classical cascade".

[Figure 2 Here]

Within Marr's original account (see figure 2a), working at level 1, the uppermost level of his three-stage hierarchy, consists of specifying the problem (e.g., the system must be able to catch balls of a particular size, thrown from a particular range of positions, within a particular range of velocities, etc.). Marr termed this the *computational* level since, in order to be solved through the design of some information processing system, the problem must be expressed in a language amenable to implementation as a computer program.

Level 2 involves specifying an *algorithm* capable of solving the problem (e.g., one algorithm might calculate the future trajectory of the ball using Newtonian mechanics, and on the basis of the ball's projected flight, calculate an intercept trajectory for the catching limb).

Finally, level 3 is the ground floor, an *implementation* of the solution algorithm (e.g., a particular LISP program, or a particular circuit board, or a particular system of mechanical devices, etc.). Marr's system of levels is hierarchical since many level-3 implementations might capture any individual level-2 algorithm, and, likewise, many level-2 algorithms might solve any individual level-1 problem. Marr (1977) gives the example of Fourier analysis, which may be calculated using one of a number of algorithms, each implemented in one of any number of different types of machine.

As an addition to Marr's original scheme, Peacocke's level 1.5 involves specifying slightly more than a description of the problem, but slightly less than a full solution algorithm (see figure 2b). At level 1.5, the designer must decide upon what Clark (1989) terms a *competence theory*, by which is meant a characterisation of the problem that is "more than merely *descriptive*" of the solution to be discovered, but is also "*suggestive* of the processing structure of a class of mechanisms" within which the solution is to be searched for (Clark 1989, p. 285). For Clark (1989, p. 285), "A competence theory, then, leads a double life. It both specifies the function to be computed, *and* it specifies the body of knowledge or information which is used by some class of algorithms." Level-1.5 descriptions are thus intended to specify the range of resources upon which the solution may draw (e.g., the space of mechanical devices from within which the tennis-ball catcher must be found).

For the manual designer, there is little difference between level 1.5 and level 2. The resources upon which the solution algorithm may draw are intimately connected with the design of the solution algorithm itself. However, for those engineers employing automatic design processes (e.g., genetic algorithms, artificial neural network learning algorithms, etc.), level 1.5 is as far down the classical cascade as it is necessary to descend.

By this stage an engineer employing an artificial neural network learning algorithm will have decided upon the number of nodes in her network, the class of learning algorithm to be employed, the format of the input to, and output from, the network, etc.; an engineer employing a genetic algorithm will have specified a mapping from genotype to phenotype, and settled upon styles of fitness appraisal, reproduction, etc. These preliminary machinations specify the resources upon which the solution may draw; they define a space of possible solutions that, it is hoped, (a) contains a viable solution, and (b) is not so large as to be overly expensive to search.

Once this has been accomplished the system will proceed, upon execution, to automatically generate a solution. The modeller has thus moved from level 1.5 directly to level 3, and is in possession of a working implementation (e.g., a system that can catch tennis balls).

For the engineer, this by-passing of level 2, which is where most of the design work takes place, is a boon. For the scientist in search of an explanation of how the system works, it is a mixed blessing as the manual move from level 1 to level 2 is also where most of the insight that fuels explanation is gained. The scientist in possession of an automatic operating model is only half-way home. She must proceed to work backwards from her level-3 implementation to a level-2 understanding of how the system actually works (e.g., faced with a working tennis-ball catcher that is implemented as an artificial neural network, effort is required in order to discover how the network achieves the task).

This characterisation of the difference between manual and automatic design offered by Clark (1990), in his terms a "methodological inversion", is reminiscent of a more general admonishment of automaticity, where the reduced effort associated with machine intelligence is compromised by the risk of a concomitant loss of understanding or insight (Bullock 2008).

For our current purposes, we use Clark's account to characterise an equivalent difference between equational and simulation modelling. Under this reading, at level 1 both equational and simulation modellers must characterise their hypothesis in an adequately rigorous manner. Each modeller then proceeds to outline the style of hypothesis testing she envisages. Here the two modellers part company. For the equational modeller, this level-1.5 account is in terms of the character of the relevant variables and their inter-relations, whilst, for the simulation builder, it is in terms of the mechanisms governing system behaviour.

The equational modeller then proceeds to struggle with her system of differential equations, until, through cunning and scholarly hard work, she arrives at a solution. In contrast, her simulation building companion merely unleashes her automatic simulation process, and waits for it to collate the (typically massive) set of data points that constitute her solution. However, upon its discovery, the equational solution is straightforwardly interpretable in terms of its components, consisting as it does of intelligible relationships between meaningful sub-parts. In contrast, the set of data-points facing the simulation-based modeller is far from straightforward to analyse, requiring considerable scholarly effort in the form of statistical analysis, sensitivity analysis, experimental manipulation, systematically ablating or excising portions of the model, testing to destruction, etc., before the implications of the data for the theory being explored are clear.

For the simulation modeller, struggling upstream from level 3 to level 2, like climbing a waterfall, is hard work, but the task can be avoided, attenuated, or postponed in a number of ways. Three popular strategies are briefly sketched below.

First, and perhaps least problematically, it can be deferred. Since unpicking a complicated simulation model is time-consuming, it can sometimes be approached piece-wise, with papers and talks presenting updates from an on-going effort to achieve full understanding of the model. This has the advantage that if things run smoothly, the job will eventually be completed, but only at the risk of mischaracterising the ongoing modelling work as a pseudo-empirical discovery quest within an artificial world rather than a purely conceptual undertaking.

Second, modellers may sometimes appeal to the inherent complexity of the system in arguing that there is nothing to be gained in pursuing a level-2 understanding of its behaviour. Bedau's (1997) definition of a macro-level emergent property of some system that we might be interested in modelling, predicting or understanding, for example, is one for which the *only* way to derive it is to let a simulation unfold and observe it arising from the simulation's micro-dynamics. Freed from the requirement to constrain models such that they are amenable to level-2 unpacking, Edmonds and Moss (2004) dispense with the widely adopted modelling maxim to 'Keep It Simple Stupid' (KISS), in favour of an alternative "anti-simplistic" 'Keep It Descriptive Stupid' (KIDS) approach that makes a virtue of the richness and descriptive complexity of their simulations without apologising for the consequent increased challenge that they pose in terms of tractability.

It is understandable that researchers convinced of the irreducible complexity of the system that they are modelling may be satisfied with a less-than-fully-analysed simulation, believing that a more principled understanding of the nature of the problem is made impossible or impractical by its complex, large-scale, emergent, chaotic, non-linear, or context sensitive nature. Although it may be the case that some problems are of this type, at what point should one be confident enough to declare that this is the nature of the beast? This issue can also be re-described, using Marr's (1977) terminology, as the difficulty in determining whether a problem has no Type-1 theory, but merely admits of a Type-2 theory.

Type-1 problems can be solved in a manner that generates insights into the nature of the problem system. For example, in solving a problem presented by the pattern of inheritance evidenced by peas, Mendel achieved some insight into the particulate nature of genetics. In doing so Mendel moved towards an *explanation* of inheritance patterns, rather than a mere *description* of them.

In contrast, problems that only have a Type-2 theory may perhaps be solved, but such solutions will not throw additional light upon the nature of the problem system. Marr offers protein folding (predicting the full structure and function of a folded protein from the linear order of its constituent amino acid monomers) as an example of a problem that (at the time he was

writing) might turn out to only have a Type-2 theory, since this problem "is solved by the simultaneous action of a considerable number of processes, *whose interaction is its own simplest description*" (Marr 1977, p. 134, Marr's italics). Marr has this to say on competition between Type-1 and Type-2 accounts:

> The principal difficulty . . . is that one can never be quite sure whether a problem has a Type-1 theory. If one is found, well and good; but failure to find one does not mean that it does not exist. . . . [T]he danger with [Type-2] theories is that they bury crucial decisions, that in the end provide the key to correct Type-1 decompositions of the problem, beneath the mound of small administrative decisions that are inevitable whenever a concrete program is designed. This phenomenon makes [such] research . . . difficult to pursue and difficult to judge.
>
> . . . With any candidate for a Type-2 theory, much greater importance is attached to the performance of the program. Since its only possible virtue might be that it works, it is interesting only if it does. (Marr 1977, p. 135).

In his final sentence, Marr presciently identifies the third strategy for dealing with simulations that are hard to understand: trading off insight and understanding for predictive accuracy. Here researchers are satisfied with a less-than-fully-analysed simulation of a real-world target system because it can be defended as a well-calibrated, empirically validated, predictive model (Oreskes 1994).

In combination, these approaches explicitly carve out a fourth strategy that Levins implicitly denies, where tractability is deliberately sacrificed for a combination of realism, precision and generality.

V

By this point we have a rather bleak account: simulation modelling of complex, poorly understood target systems can tempt modellers to build systems that are beyond their comprehension in a misguided effort at combining generality, precision and realism. An illusion of tractability is created by the powerful automaticity of computational tools. But there is no free lunch, in that what is won on the swings of easy construction is lost on the roundabouts of intractable analysis. In this situation simulation modellers make do with systems that are to some extent general, precise and realistic, but not fully understood. Consequently such simulations can

come to resemble the mysterious real world rather than a traditional scientific model, lending an experimental, empirical flavour to their exploration.

Since experimentation is a legitimate scientific activity, simulation communities can come to treat overly complicated simulation models as legitimate objects of enquiry—understanding them becomes an end in itself. By contrast, some communities may replace efforts towards understanding a model with efforts towards predictive accuracy, becoming more interested in hind-casting, empirical validity, etc., rather than achieving an understanding of the reasons for the model's behaviour.

However, while there are clear pathologies at work here, there is something to be salvaged for artificial worlds by noting that the divide between tractable and intractable is not set in stone. It is not the case, as Orzack and Sober (1993) suggest, that limits on model tractability are biologically determined by the unchanging capacity of our *Homo sapiens* brains. Improved physical or mechanical power or uncovering new paths and passes can bring previously unscalable heights within reach. Likewise, improved simplifying assumptions, theoretical frameworks, representational re-descriptions, organising concepts, etc., can simplify or unlock previously intractable problems, bringing them within reach of existing analytical tools or stimulating the creation of effective new ones.

[Figure 3 Here]

It is by operating at or around a current tractability ceiling that the nature of this ceiling will be examined closely and changed (see figure 3). Barriers to achieving a more theoretically secure position with respect to some problem will tend to be overcome as a consequence of new insights and tools, conceptual frameworks, etc. Working with models that are challenging in their tractability is a good way to identify the weak or illusory parts of a tractability ceiling in order to push beyond it. By analogy with Vygotsky's (1978) "zone of proximal development", simulation artefacts can act to increase the modeller's "zone of proximal theoretical development", the area of theoretical development that the modeller cannot access alone, but can access with the assistance of an external agency—in this case an automated model. This account is in part simply a reassertion and reinforcement of a familiar take on simulations, casting them as tools for thinking (Di Paolo, Noble & Bullock, 2000)—but in this case we are stressing a special kind of thinking that re-examines theoretical commitments.

Several examples of simulation at tractability ceilings are to be found at the intersection between computing and biology. This interface has a long and interesting history with significant contributions from pioneers of computing science such as von Neumann, Turing and Babbage

(Bullock 2008). More recently, John Maynard Smith and William Hamilton, both pioneers of mathematical modelling in evolutionary biology, have made low-key use of computer simulations in developing new mathematical treatments of evolutionary problems.

One of Maynard Smith's most significant contributions to evolutionary biology was the origination of the Evolutionary Stable Strategy (ESS) concept and the development of the game theoretic mathematical apparatus for identifying ESSs (Maynard Smith 1982). However, his first published ESS model, jointly authored with Price (Maynard Smith & Price 1973), was solved via a simple *computational* model. Within a year, he was able to analytically treat similar models mathematically (Maynard Smith 1974) and thereafter resorted to computational approaches in his published papers only occasionally. However, at the end of his career he was still working with simple simulation codes and attempting to get them reformulated into more modern programming languages in order that they would run on modern machines.

Hamilton is also associated almost exclusively with mathematical models within evolutionary biology, but presented a series of unpublished simple simulation models at a meeting on the simulation of adaptive behaviour in the UK in 1994[2]. He had never thought to publish these models as they had merely served to unlock progress towards mathematical models that could be published in their stead.

Sober (1996) summarises this use of simulation somewhat dismissively as enabling theoreticians to "get a feel for the model's dynamics" in cases for which the model proves to be unsuited to analytic methods. This familiar but powerful role for simulation is under-appreciated, largely untaught and informal, but is perhaps core to the use of simulation more generally, and has the potential to teach us much about the interaction between the context of model discovery and model justification.

VI

The lure of functional models that combine elements of precision, generality and realism at the expense of tractability is felt particularly strongly by simulation modellers because the automaticity of their models has the effect of decoupling two traditionally tightly coupled elements of model tractability: model synthesis and model analysis. This decoupling allows the construction of "artificial worlds" that are operationally interesting without necessarily being

[2] The Third International Conference on the Simulation of Adaptive Behavior, Brighton, UK, August 8-12, 1994.

fully understood. As such, and contra some simulation practice, artificial worlds are artefacts that demand considerable post-construction analytic effort.

However, working at and around such "tractability ceilings" is the right way to achieve insights that unlock new and better models and theories, and by projecting us through such ceilings, albeit temporarily, even a less-than-fully understood simulation model can give us, not a window on a new artificial world of empirical facts and findings, but a new perspective on an existing world of ideas, assumptions, commitments and questions.

Acknowledgements

The content of this paper has benefitted from reviewer comments and discussion with colleagues and students at the University of Leeds and University of Southampton, and with the participants and audiences at several research seminars, including those at the University of Sussex (Informatics), University of California San Diego (Science Studies), and London School of Economics (Philosophy of Science), and meetings including the third "Workshop on Epistemological Perspectives on Simulation" (Lisbon, Portugal 2008), "Thought Experiments and Computer Simulations" (Paris 2010), and "Knowing and Understanding Through Computer Simulations" (Paris 2011).

References

Barandiaran, X. and Moreno, A. 2006. "ALife models as epistemic artefacts," in Rocha, L.M., Yaeger, L.S., Bedau, M.A., Floreano, D., Goldstone, R.L., and Vespignani, A., eds., *Proceedings of the Tenth International Conference on Artificial Life*, pages 513–519, Cambridge, MA: MIT Press

Bedau, M.A. 1997. "Weak emergence," in Tomberlin, J., ed., *Philosophical Perspectives: Mind, Causation, and World*, pages 375–399, Oxford: Blackwell.

———. 1998. "Philosophical content and method of artificial life," in Bynum, T.W. and Moor, J.H., eds., *The Digital Phoenix: How Computers Are Changing Philosophy*, pages 135–152, Oxford: Blackwell.

Bedau, M.A., McCaskill, J.S., Packard, N.H., Adami, S.R.C., Green, D.G., Ikegami, T., Kaneko, K., and Ray, T.S. 2000. "Open problems in artificial life," *Artificial Life*, 6 (4): 363–376.

Braitenberg, V. 1984. *Vehicles*, Cambridge, MA: MIT Press

Bullock, S. 2008. "Charles Babbage and the emergence of automated reason," in Husbands, P., Holland, O., and Wheeler, M., eds., *The Mechanical Mind in History*, pages 19–39, Cambridge, MA: MIT Press

Casti, J.L. 1997. *Would-be Worlds: How Simulation is Changing the Frontiers of Science*, New York, NY: John Wiley.

Clark, A. 1990. "Connectionism, competence and explanation," in Boden, M.A., ed., *The Philosophy of Artificial Intelligence*, pages 281–308, Oxford: Oxford University Press.

———. 1989. *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing*, Cambridge, MA: MIT Press

Deneubourg, J.-L., Theraulaz, G., and Beckers, R. 1991. "Swarm-made architectures," in Varela, F. and Bourgine, P., eds., *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, pages 123–133, Cambridge, MA: MIT Press

Di Paolo, E., Noble, J., and Bullock, S. 2000. "Simulation models as opaque thought experiments," in Bedau, M.A., McCaskill, J.S., Packard, N., and Rasmussen, S., eds., *Proceedings of the Seventh International Conference on Artificial Life*, pages 497–506, Cambridge, MA: MIT Press

Edmonds, B. and Moss, S. 2004. "From KISS to KIDS—an 'antisimplistic' modelling approach," in Davidsson, P., Logan, B., and Takadama, K., eds., *Multi-Agent and Multi-Agent-Based Simulation*, pages 130–144. Berlin: Springer.

Gould, S.J. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*, New York, NY: W. W. Norton & Co.

Hamilton, W.D. 1971. "Geometry for the selfish herd," *Journal of Theoretical Biology*, 31: 295–311.

Kitano, H., Hamahashi, S., Kitazawa, J., Takao, K., and Imai, S. 1997. "The virtual biology laboratories: A new approach of computational biology," in Husbands, P. and Harvey, I., eds., *Proceedings of the Fourth European Conference on Artificial Life*, pages 274–283, Cambridge, MA: MIT Press

Langton, C. 1989, "Artificial Life," in Langton, C. ed., *Artificial Life*, pages 1–47, Redwood City, CA: Addison-Wesley.

Lenski, R.E. 2004. "The future of evolutionary biology," *Ludus Vitalis*, 12: 67–89.

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

Marr, D. 1977. "Artificial intelligence—a personal view," in Boden, M.A., ed., *The Philosophy of Artificial Intelligence*, pages 133–147, Oxford: Oxford University Press. Collection published in 1990.

Matthewson, J. and Weisberg, M. 2009. "The structure of tradeoffs in model building," *Synthese*, 170 (1): 169–190.

Maynard Smith, J. 1974. "The theory of games and the evolution of animal conflicts," *Journal of Theoretical Biology*, 47: 209–221.

———. 1982. *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.

Maynard Smith, J. and Price, G.R. 1973. "The logic of animal conflict," *Nature*, 246: 15–18.

Odenbaugh, J. 2003. "Complex systems, trade-offs and mathematical modeling: A response to Sober and Orzack," *Philosophy of Science*, 70: 1496–1507.

———. 2006. "The strategy of "The strategy of model building in population biology"," *Biology and Philosophy*, 21: 607–621.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. 1994. "Verification, validation, and confirmation of numerical models in the earth sciences," *Science*, 263 (5147): 641–646.

Orzack, S.H. and Sober, E. 1993. "A critical assessment of Levins's "The strategy of model building in population biology"," *The Quarterly Review of Biology*, 68: 533–546.

Peacocke, C. 1986. “Explanation in computational psychology: Language, perception and level 1.5,” *Mind and Language*, 1 (2): 101–123.

Peck, S.L. 2004. “Simulation as experiment: A philosophical reassessment for biological modelling,” *Trends in Ecology & Evolution*, 19: 530–534.

Premo, L.S. 2006. “Agent-based models as behavioral laboratories for evolutionary anthropological research,” *Arizona Anthropologist*, 17: 91–113.

Ray, T.S. 1994. “An evolutionary approach to synthetic biology: Zen and the art of creating life,” *Artificial Life*, 1: 179–209.

Silverman, E. and Bullock, S. 2004. “Empiricism in artificial life,” in Pollack, J., Bedau, M.A., Husbands, P., Ikegami, T., and Watson, R.A., eds., *Proceedings of the Ninth International Conference on Artificial Life*, pages 534–539, Cambridge, MA: MIT Press

Sober, E. 1996. “Learning from functionalism—Prospects for strong artificial life,” in Boden, M.A., ed., *The Philosophy of Artificial Life*, Oxford: Oxford University Press.

Tucker, V.A. 1993. “Gliding birds: Reduction of induced drag by wing tip slots between the primary feathers,” *Journal of Experimental Biology*, 180: 285–310.

Vygotsky, L.S. 1978. *Mind in Society: The Development of Higher Psychological Processes*, Cambridge MA: Harvard University Press.

Watson, R.A. 2006. *Compositional Evolution: The Impact of Sex, Symbiosis and Modularity on the Gradualist Framework of Evolution*, Cambridge, MA: MIT Press

Watts, D.J. and Strogatz, S.H. 1998. “Collective dynamics of ‘small-world’ networks,” *Nature*, 393 (6684): 440–442.

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

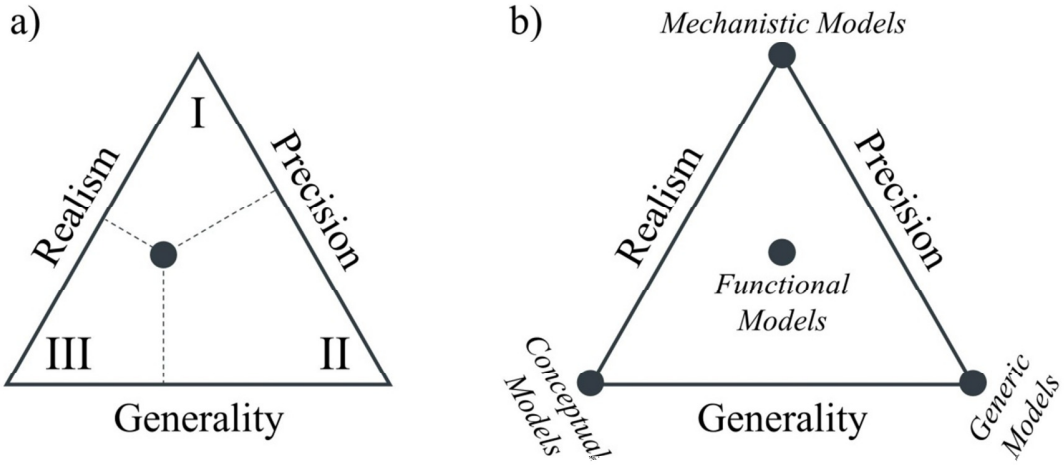

Figure 1: (a) Diagrammatic representation of Levins' (1966) three modelling strategies, I, II, and III, each of which sacrifices one attractive model property in order to achieve two others; (b) These strategies can be relabeled according to work by Barandiaran and Moreno (2006), who introduce a fourth category: functional models that combine elements of realism, precision and generality.

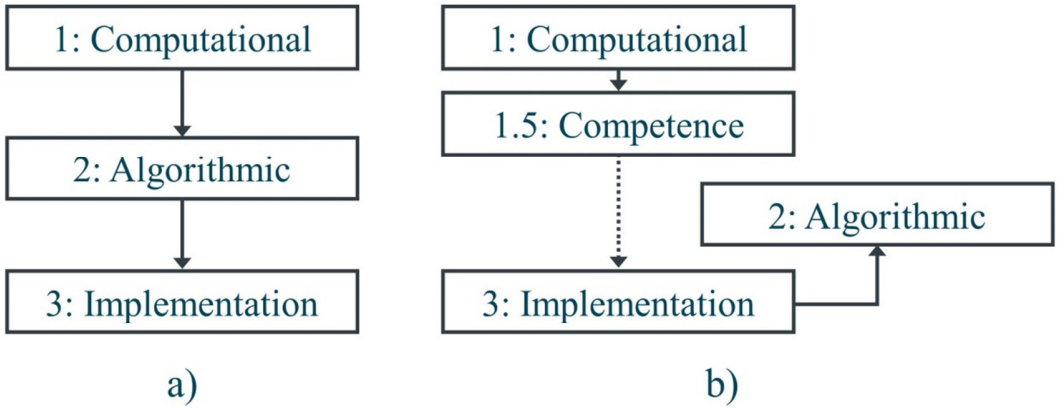

Figure 2: (a) The three hierarchical levels of description invoked in Marr's (1977) "classical cascade". Arrows indicate the direction of travel for model construction in cognitive science; (b) Clark's (1990) reworking of this scheme to include Peacock's (1986) "Level 1.5" and the automatic move from this level to a working implementation (dashed arrow), followed by the "inverted" move required to manually reconstruct an algorithmic level explanation.

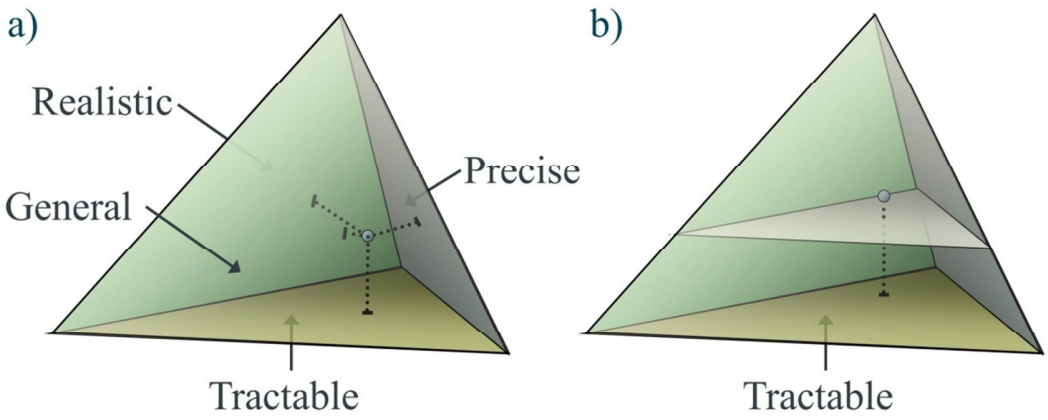

Figure 3: (a) A four-dimensional Levinsian space of modelling activity. Here the closer a point is to a tetrahedron face, the more of the property associated with that face it possesses. Hence the floor of the space corresponds to maximally tractable models; (b) A simulation model existing above a “tractability ceiling”.