Animals versus animats: or why not model the real iguana?

Running head: Animals vs. Animats

Barbara Webb

Institute for Perception, Action and Behaviour, School of Informatics,
University of Edinburgh

10 Crichton Street, Edinburgh EH8 9AB, U.K.
bwebb@inf.ed.ac.uk
phone: +44 131 651 3453
fax: +44 131 650 6899

Abstract

The overlapping fields of ‘Adaptive Behaviour’ and ‘Artificial Life’ are often described as novel approaches to biology. They focus attention on bottom-up explanations, and how lifelike phenomena can result from relatively simple systems interacting dynamically with their environments. They are also characterised by the use of synthetic methodologies, that is, building artificial systems as a means of exploring these ideas. Two differing approaches can be distinguished: building models of specific animal systems and assessing them within complete behaviour-environments loops; and exploring the behaviour of invented artificial animals, often called ‘animats’, under similar conditions. An obvious question about the latter approach is: how can we learn about real biology from simulation of non-existent animals? In this paper I will argue, first, that animat research, to the extent that it is relevant to biology, should also be considered as model building. Animat simulations do, implicitly, represent hypotheses about, and should be evaluated by comparison to, animals. Casting this research in terms of ‘invented agents’ serves only to limit the ability to draw useful conclusions from it, by deflecting or deferring any serious comparisons of the model mechanisms and results with real biological systems. Claims that animat models are meant to be ‘existence proofs’, ‘idealisations’, or represent ‘general’ problems in biology do not make these models qualitatively different from more conventional models of specific animals, nor undermine the ultimate requirement to justify this work by making concrete comparisons with empirical data. It is thus suggested that we will learn more by choosing real, and not made-up, targets for our models.

Keywords: models, synthetic, simulation, simple animals, minimal cognition.

1. Introduction

The issue of how animals achieve adaptive behaviour has been part of the study of biology throughout its history. The relatively recent fields of ‘Adaptive Behaviour’ (AB) and ‘Artificial Life’ (ALife) are distinguished from standard biological sciences by the use of synthetic methodologies to investigate this problem, that is, by building systems that exhibit the behaviour of interest. In particular, synthesis is often used with the aim of demonstrating how some high level phenomena of life, adaptive behaviour or intelligence can emerge from low level interactions of relatively simple components. As described by Langton (1989), ALife:

> *“complements the traditional biological sciences concerned with analysis of living organisms by attempting to synthesize life-like behaviors within computers and other artificial media... The field as a whole represents an attempt to vastly increase the role of synthesis in the study of biological phenomena”.*

Guillot and Meyer (1997) write of the AB approach that:

> *“the synthesis of animats can be expected to help in understanding how real animals manage to survive in real worlds”.*

Understanding biology is not necessarily the only aim of work in these fields, but significant claims of biological relevance are often made. The main question I wish to explore here is whether and how “the role of synthesis” differs from the role of modelling in biology. Have we really developed a new methodology for understanding real life?

To focus this discussion, I will draw examples mostly from one strand of this research, where the questions addressed are how particular sensorimotor capabilities of organisms or agents support interesting, potentially intelligent behaviour. Braitenberg’s 1984 book ‘Vehicles’ is one of the key inspirations. He describes a series of thought experiments in which increasingly intelligent (and life-like) vehicles are constructed, evolve and learn. At each stage, it is argued, surprisingly complex external behaviour will result from apparently simple internal structures. Another important stimulus in the field was the robots built in the 1980s by Brooks and his colleagues (reviewed in Brooks, 1990). They demonstrated, in real devices rather than mere thought experiments, how having the right physical interfaces and relatively simple control mechanisms (eschewing the internal world modelling approaches of traditional AI robotics) could produce successful behaviours in real environmental contexts. W. Grey Walter's robots (1950,1951) are rightly seen as early forerunners of this work. These approaches meshed naturally with existing ideas from ecological psychology (Gibson, 1979) that were re-emerging in AI and cognitive science (e.g. Clancy, 1997) as an interest in active, situated and embodied aspects of cognition (e.g. Clark, 1997; Pfeifer & Scheier, 2000). However, the new AB and ALife influence could be seen in the suggestion that investigation of these issues should start with relatively simple, yet complete, systems (such as insect-like reactive behaviours) and approach more complex capabilities (such as human intelligence) only incrementally.[1]

One obvious way to proceed towards these goals is to study real insects (or other 'simple' biological systems) and to try to model them in a way that encompasses sensing and action, bottom-up control, embodiment, environmental interaction, and so on. An early example, which anticipated many of the above principles, is Arbib's work on modelling prey capture in toads (Arbib, 1982). My own work on robotic implementations of cricket phonotaxis (e.g. Webb, 1995) is another example, which I will describe more thoroughly below. I have reviewed similar work by other researchers in Webb (2000) and have argued that this is, essentially, a methodology of model building; and can thus usefully be compared to other kinds of biological modelling on a number of dimensions (Webb, 2001).

However another approach, common in AB and ALife research, follows instead the suggestion made by Dennett (1978):

> *"one does not want to get bogged down with technical problems in modeling the cognitive eccentricities of turtles if the point of the exercise is to uncover very general, very abstract principles… So why not then make up a whole cognitive creature, a Martian three-wheeled iguana, say, and an environmental niche for it to cope with?" (p.104).*

I will refer to this as the 'animat' approach – the invention and study of an artificial creature that does not correspond to any specific real animal, but is (nevertheless) intended to provide some insight into real issues in biological or cognitive science. The term 'animat' was first coined by Wilson (1985) who used an agent moving on a 18x58 hexagonal grid, able to distinguish three cell states (labelled 'food','tree' and 'empty'), to explore how classifier systems might be used to learn strategies for efficient gathering of 'food' and avoidance of 'trees'. In Wilson (1991) he further describes this approach to *"the scientific understanding of intelligence, or how mind arises from brain"* as *"simulating and understanding complete animal-like systems at a simple level"* (p.15). A more recent example is the work of Beer (e.g. Beer, 1996) on evolving and analysing "minimally cognitive creatures", which I will describe in more detail below. There is again much similar work;

---

[1] *"In building artificial creatures we might well make progress by starting with low expectations and gradually building upon our experience…One approach then, is to aim initially for the intelligence level of lower animals (e.g. insects) and respect the constraints that biology seems to have worked under."* (Brooks, 1989). Brooks in fact did not follow his own advice in this regard, jumping in one step from insect-like robots to a humanoid robot project (Brooks and Stein, 1994).

many relevant examples can be found in the proceedings of *ALife* and *Simulation of Adaptive Behaviour* conferences, and in the journals *Artificial Life* and *Adaptive Behaviour*.[2]

In what follows I will argue that this approach, insofar as it is intended to be relevant to biology, should also be viewed as modelling; but that the focus on the invention of 'made-up' target creatures is largely counter-productive because it makes it difficult to verify or evaluate the results. The presentation of such work in terms of 'existence proofs', 'alternative forms of life', 'conceptual clarification', 'idealisations' and so on does not suffice to side-step this fundamental limitation. To make the comparison concrete I will first describe in more detail two specific examples of 'animal modelling' and 'animat modelling', already mentioned above, that is Webb's 'cricket robot' and Beer's 'minimally cognitive agent'.

2. Two research approaches

2.1 Modelling adaptive behaviour in animals

In research over more than ten years (e.g. Webb 1995, Lund *et al.* 1997, 1998, Webb & Scutt 2000, Reeve and Webb 2003, Reeve *et al.* 2005) we have been modelling the phonotaxis behaviour of crickets. Female crickets can recognise and locate the calling songs of male crickets. This target system was chosen for study due to the extensive behavioural and physiological data (an early review is Huber & Thorson 1985; more recent research is reviewed in Hennig *et al.* 2004) that could be fed into a model, and used to evaluate it. It was also plausible to investigate this behaviour as a complete loop from sensing to action, addressed at the level of single neurons and their connectivity. The behaviour depends on a well-tuned physical auditory apparatus. The two ears, on the cricket's forelegs, are linked by a tracheal tube, thus acting as pressure difference receivers to detect the direction of sounds of a specific frequency (Michelsen *et al.*, 1994). There are also interesting emergent properties such as apparent 'choice' between songs that may be a simple consequence of interaction of the localisation mechanism and the complex sound field produced by multiple males singing. Phonotaxis behaviour interacts in interesting ways with other sensorimotor loops, such as the optomotor reflex (Böhm *et al.*, 1991), allowing the model to be gradually extended (Webb & Harrison, 2000) towards more complex capabilities.

The more recent models we have built simulate spiking neuron activity using a leaky-integrate-and-fire representation (based on the approach described by Koch, 1999). Synaptic input is modelled as a change in conductance with a specific time course; the synaptic 'weight' is a function of the battery potential (positive for excitation, negative for inhibition), the size of conductance change

---

[2] 'Animat' research can sometimes overlap with the field of 'agent' research but I am referring here only to 'agent' simulations that claim to be addressing problems relevant to biology, in particular, about the mechanisms of adaptive behaviour in real animals.

occuring for each spike, and the time constant of conductance decay. This is a more realistic model than the more common treatment in integrate-and-fire neural modelling of synaptic input as an instaneous injection of current. The synapses also have some temporal dynamic properties, depression (decreased conductance with repeated spikes) and facilitation (increased conductance with repeated spikes), each with their own time course. However there is no permanent adaptation of weights (no learning). The networks consist of 4-20 units, and are generally hand-designed and hand-tuned, although we have explored the use of genetic algorithms to tune some parameters. Parts of the network are based on identified neurons in the cricket, for example the thoracic auditory interneurons known as AN1 and ON1 (Wohlers and Huber, 1982). Other parts are more speculative, either proposing specific connectivity between neurons that have been identified but not anatomically linked, e.g. the brain neurons BNC1 and BNC2 (Schildberger 1984), or proposing new neural elements e.g. a ‘fast’ neuron that – based on recent cricket behavioural data (Hedwig & Poulet, 2004; Poulet & Hedwig, 2005) – may be involved in a rapid reflex response to sound pulses that is modulated by a slower recognition process.

These neural simulations have in most cases been tested in the context of a robot implementation. The robots (several different platforms have been used) have an analogue electronic auditory system to represent the physical auditory receptor mechanism of the cricket. The auditory input is translated into spike trains for the neural simulator, and the output of the simulator is used to drive the wheels (or in one case, ‘whegs’ – a hybrid wheel-leg mechanism (Horchler *et al.,* 2004)) of the robot. Thus the system can be tested in real experiments, using the same stimulus that is used for real crickets, to evaluate whether comparable behaviour can be produced.

It is worth noting at this juncture that this work has not simply been a case of using the robot model to verify or falsify existing hypotheses in the biological literature. It has involved substantial integration of disparate information about auditory mechanisms, neural data, and behavioural observations. It has led to the proposal of several novel hypotheses about the function, some of which have been supported, and some contradicted, in subsequent investigations. It has also made much more apparent the areas in which biological data is most critically lacking. One example is that the detailed dynamics of the cricket’s response (how quickly it turns, the time constant of integration of signals into the response) has only recently been studied (Hedwig & Poulet, 2005) and there is still almost no information about the precise leg movements during phonotactic tracking that would be needed for a walking robot implementation.

There is much similar work that I would classify as part of the ‘animal modelling’ approach, some of which I have described in previous reviews (Webb 2000, 2001). To give a few recent examples:

- Human walking behaviour is a subject that has been widely investigated using simulated and real robots (though many 'humanoid' robots do not have human-like walking). One interesting strand of this research is the investigation of passive walking machines. In these, the mechanics of human walking are represented in physical devices that mimic the hip and knee joints, the kneecap as an endstop, the foot, and sometimes the counter swing of the arms e.g. the work of Collins *et al.*, 2005 who say "We study human locomotion by designing and testing walking machines that we compare to humans in terms of morphology, gait appearance, energy use, and control" (p.1082). It has been demonstrated that convincingly human-like walking gaits emerge from the dynamics of these simple physical systems.
- The larvae collecting behaviour of the ant *Leptothorax albipennis* has been modelled using simulated and robot agents, to see which minimal set of action rules both most efficiently and most accurately reproduce the results. This can be used to infer what variables the ants themselves may be using to perform the task (Scholes *et al.,* 2004).
- The 'Psikharpax' project described in Meyer *et al.,* (2005) includes simulations of rat brain mechanisms for navigation (based on hippocampal place-cell data), integrated with action selection (based on basal ganglia-thalamus-cortex loops) and learning (based on dopamine reinforcement mechanisms) and aims to implement these on a robot with appropriate sensorimotor capabilities with the aim to "better understand the control mechanisms of rats" (p.221).. However it should be noted that Meyer *et al.* use the term 'animat' in a way that would include this kind of model – "*animats,* i.e., simulated animals or real robots whose sensors, actuators and control architectures are as closely inspired from those of animals as possible" (p.211) – whereas I am here limiting its application to invented creatures as described next.

2.2 Investigating adaptive behaviour in animats

Over a similar time period as the cricket robot research described above, Beer and colleagues have been carrying out an investigation into "minimally cognitive agents" using simulated animats (Beer, 1996, 1997, Slocum *et al.,* 2000, Beer 2003a, Goldenberg *et al.*, 2004). The methodology is discussed in detail in Beer (2003a, 2003b) with a particular focus on how the agent behaviour can be analysed using methods from dynamical systems theory. The main task explored is the ability of an agent to actively categorise different stimuli, by moving to intercept items of one shape and to avoid items of another shape. Beer provides an explicitly (though perhaps not exclusively) biological motivation for this work as a "scientific approach to the neural basis of behaviour" (p.239). He is interested in discovering how small neural circuits can produce adaptive sensorimotor functioning, in the context of complete sensorimotor loops through the environment, and how this can be explained using a dynamical framework rather than a computational one.

The agents in these simulations have small continuous time recurrent neural networks (CTRNNs). The neural units are modelled as continuously varying values (interpretable as equivalent to the firing rates of real neurons) that decay with a set time constant. Synaptic input is the sum of the weighted activity, after passing through a logistic function, of all connected neurons, plus any external input. For the system described in Beer (2003a) the network has a fixed architecture consisting of 7 input neurons projecting to 5 fully interconnected interneurons, which project to two motor neurons. The weights of all the connections are determined using an evolutionary algorithm, evaluating each agent over a set of trials by its ability to minimise its final distance to one shape, and maximise the distance to the other.

This neural system is embedded in a simulated agent, which has seven sensors, evenly distributed across a $\pi/6$ 'field of view'. If an object intersects the axis of a sensor, the sensory neuron is activated by an amount inversely proportional to the distance of the intersection from the agent. The sum of the activity of the two motor neurons determines the horizontal velocity of the agent. Objects fall from above, with a fixed velocity, and a random horizontal offset from the agent's starting position. A trial ends when the object hits or passes the agent. Agents that have successfully evolved to perform the task can be tested in experiments (e.g. with stimuli varied systematically from that used in the evolutionary process) to evaluate exactly how the agent is able to produce the observed behaviour. An important conclusion drawn is the need to understand this behaviour as emerging from the continuous interaction of neural, bodily and environmental factors, rather than treating it as a computational transformation of sensory input to motor output. To this end, dynamical analyses are applied at several levels: to the overall behaviour, to specific stimuli interactions, and to the functional role of specific neurons (Beer, 2003a).

It should be evident that this integrates a number of issues in sensing, neural circuits, adaptation and motor control. It proposes a rather novel hypothesis about the nature of cognitive capabilities, i.e., that they involve such a tight interlinking of neuron-neuron, neuron-body, and body-environment dynamics that the standard kinds of computational decompositions used in investigating cognitive systems may be simply false. It also points out where there are substantial gaps in the methodological tools available for analysing such systems, and attempts to fill some of these gaps. Given that in what follows I take a rather critical view of this work, I should stress that this example was chosen due to the fact that within the field of animat modelling, it represents some of the most thorough, clearly motivated, and well-executed research; it has been quite influential in the field;

and it is convenient for developing my argument that the underlying methodological assumptions have been so lucidly expounded by Beer.

Some other examples that illustrate the scope of the animat approach include:

- Investigation of the behavioural consequences of evolved morphologies, e.g. to determine if there is an adaptive advantage to body symmetry for locomotion (Bongard and Paul, 2000). In this case the agents are simulated using a physics engine and consist of identical spherical units that can be connected to each other by hinge joints in six cardinal directions. The behaviour is controlled by a recurrent network of sigmoidal neurons that is evolved in concert with the morphology. The higher efficiency of the symmetric systems is taken to support the possibility that efficiency was a factor in the evolution of bilateral symmetry in biological systems.
- Study of the origins and evolution of communication and language (e.g. MacLennan 1992, MacLennan and Burghardt 1993). This involves a population of simulated agents able to have simple linguistic interactions, genetic interactions, and sensorimotor interactions with an environment which contains various entities that they come to label in a common manner. These systems are then studied for effects such as population size and distribution on the development of vocabulary and syntax. MacLennan and Burghardt (1993) describe this work as “synthetic ethology” in which “the goal is to use the synthetic approach to understand the natural phenomena” (p.163).
- The ‘synthetic epistemology’ approach explored in the Distributed Adaptive Control (DAC) architecture by Verschure and colleagues (e.g. Verschure et al, 2003). This falls somewhat closer to the 'animal modelling' end of the continuum than the previous examples, in that it utilises models of learning based more directly on biological examples (e.g. classical conditioning using a Hebb rule), and some analogies to vertebrate brain structures are described. But the model *“is not inspired by the anatomy, physiology or neuropsychology of hippocampus and cortex, but addresses the general problem of communication between different neural structures in the context of behaviorally realistic tasks and well-evaluated models of learning”* (Voegtlin and Verschure, 1999 (reprint 34/55). The architecture is tested in an animat, which is not designed to represent any specific animal, which performs ‘foraging tasks’ to locate targets and avoid obstacles. The behaviour is assessed in terms of efficiency and stability of learning, as aspects of the architecture are varied, rather than compared directly to learning or foraging data from biology.

2.3 Similarities

It should be evident from the above descriptions that there are many ways in which the two main examples I have described – my 'cricket robot' and Beer's 'minimally cognitive agent' – are similar. The behavioural competence studied is closely comparable: both Webb's robot, and Beer's agent have to discriminate the correct stimulus (cricket songs with the right temporal pattern, items of a particular shape) from the wrong stimulus (other sounds, an alternative shape) and move towards the former. Both attempt to account for behaviour at the same level of description, that is, examining the functional roles of individual neurons in small circuits for an agent in a particular action-environment context.

Although the models have been embodied in different mediums (robot vs. computer simulation), in the current context I do *not* regard this as a significant difference. It is not difficult to envisage a robotic implementation of Beer's agent. Indeed, in work closely related to that of Beer, Husbands et al (1995) used a gantry robot to evolve a small neural network to control an agent that would approach triangles and not squares, and carried out a dynamical analysis of the behaviour. The animat work by Verschure *et al.* (2003), mentioned above, has been implemented both in simulations and real robots. In a complementary fashion, virtual environments are often used for simulation of real animals, e.g. the work by Neumann & Bulthoff (2002) on visual control of motion in the fly. In Neumann *et al*. (2001) they argue that the experiments possible in simulation are in fact more realistic for representing the animal than any plausible robotic implementation could be with current technology.

So what are the key distinguishing factors between the approaches? The most obvious difference is that Beer's agent is described as an idealised model of a generic cognitive agent; whereas the cricket robot attempts to be a realistic model of a specific animal. But what do 'idealised' and 'realistic' actually mean in this context? How much does an idealised model differ in detail or accuracy from a realistic one? What, if anything, is more 'general' about Beer's agent categorising shapes than the cricket robot recognising songs? Are idealisations models at all? I will try to explore these and related questions in what follows, and in the process it is hoped also illuminate some issues relating to the nature of scientific modelling in general.

3. Are animats models?

I have argued previously (Webb 2001, 2006) that the cricket robot (and 'biorobots' in general) are best viewed and evaluated as models. In this section I will argue that Beer's agent (and animats in general) should also be viewed and evaluated as models. But first, as the term 'model' can be used in a number of different ways, my intended meaning should be clarified by reference to the following scheme of scientific reasoning (c.f. figure 1 in Webb 2001).

A *target* phenomenon in the world is identified – where a specific target might be considered a ‘model system’ for a wider class of phenomena. For example, the cricket is considered a ‘model animal’ for investigating auditory behaviour. A hypothesis or causal account is then offered as an explanation of the phenomenon – this may be called a ‘theoretical model’ - which represents the target in the same way that any language description represents a state of affairs (see also Callendar and Cohen, 2005) although the ‘language’ used may be more precise than natural language. For example, we hypothesise that a particular neural circuit is sufficient to produce both recognition and orientation in the cricket. The hypothesis may draw on an analogy to some other, presumably better understood, system – often called a ‘source’ model. Our proposed neural circuit was partly inspired by the known temporal filtering properties of short-term synaptic dynamics in other sensory systems. The extent to which the hypothesis accounts for existing data and predicts new data from observations on the target phenomenon is taken to support its status as an explanation[3]. These observations may include both the overt behaviour of the system and any structural components and relations that are revealed by ‘dissection’ of the system. For example, our proposed neural circuit predicts that certain song preferences should be shown by the cricket, and that a particular neural connectivity should be found.

For a moderately complex hypothesis, demonstrating what data it can account for and will predict may require implementation in a mathematical, computational or even physical model – and *this* is the specific meaning of ‘model’ that applies to the cricket robot. Such a model represents (usually in some different medium) the mechanisms and causes described in the hypothesis, and can be solved, run or tested to see what their consequences are[4]. Thus the cricket robot is a mixed physical and computational implementation of a particular hypothesis about an observed phenomenon of sensorimotor behaviour in a certain animal. It is relevant to biology to the extent that its components and behaviour can be directly compared to that animal. Note, however, this is referring to the range of *possible* comparisons, not necessarily the *successful* matching of parts or results, as will be further discussed below. The robot does not cease to be a model if it only partially or approximately explains, or even if it largely fails to explain, cricket phonotaxis, though it might for these reasons be considered a poor model. It is a model by reason of its intended use.

---

[3] Other factors, such as simplicity, or coherence with other theories, may also enter into determining this status, in particular if there exist alternative hypotheses also able to account for the data.

[4] It is worth noting that the two stages, proposing a hypothesis, and implementing that hypothesis as a model, are rarely so distinct in practice as this description implies.

Do animats, such as Beer's minimally cognitive agent, fit into this schema, and if so, how? A number of different views on this issue can be distinguished.

4.1 Bio-inspired engineering

Some research involving reference to animals and the construction of artificial systems that appears similar (e.g. my cricket robot and the sound localising robot described in Andersson *et al.* 2004) differs significantly in intent (to explain a biological system vs. solving a technical problem using some ideas from biology). 'Biological inspiration' differs from biological modelling because the animal is not the *target* of the model building but rather the *source* model or analogy. The target is instead some technical problem, and the hypothesis is a proposed solution to that problem that draws on what is known about the biological system; this solution is often tested by implementation in a mathematical, computational or physical model. Another example would be the use of 'ant colony' algorithms as a bio-inspired solution for communication routing problems (Dorigo and Stützle, 2004) compared to biological modelling of ant colony behaviour e.g. Calenbuhr and Deneubourg (1990). If the intent is to engineer a successful artificial system, then the relevance of the results to biology is not a criteria for judging success, and failure to compare the components and behaviour of the model to some real animal target is not a valid criticism. However, although some research in ALife and Adaptive Behaviour has an engineering purpose (e.g. examples are reviewed in Kim and Cho, 2006), the basic aims as described by Langton, Meyer and Wilson, Beer and others are not technological but scientific. For the purposes of this paper, it will be assumed that animat research is intended to be relevant to biology, and is not just bio-inspired engineering.

4.2 'Pure' exploration

A slightly different view to the practical engineering perspective is that animats are simply worth exploring in their own right, as analytical systems, more or less based on biological principles that have interesting properties. Much work in evolutionary algorithms has this character, exploring the consequences of systems developing via the evolutionary principle of replication with variation, without further concern as to how the results relate to natural evolution. Such work *may* turn out to be relevant to biology, but interpretation of the systems as models that represent a particular biological problem is not considered necessary to justify the ongoing research programme. Taylor (1989) describes similar models in ecology as "exploratory tools" that do not need to be justified by either fitting to data or plausibility of their assumptions:

> *"It can be explored systematically as a mathematical system, e.g. how does the system's behaviour change as its parameters change or become variables, as time lags are added, and so on? Such mathematical investigations may help us derive new questions to ask, new terms to employ, or different models to construct." (p.122)*

However, he notes:

> *"Strictly speaking, without a quantitative analysis of correspondence [to observations], the insights from exploration are insights about a mathematical system. Their relevance to biology is yet-to-be-established; truth or falsity is a moot point…Qualitative insights might have misguided research. The categories of exploratory models, often chosen with an eye to mathematical tractability, may have obscured profound issues about biology."(p.123-124)*

Whether such exploration is science is debatable (is pure mathematics science?) but it seems reasonable to say that until or unless these explorations are used to make empirical claims about real life they are not biological science. Although a strong flavour of this 'pure' exploratory attitude is found in much animat research, in most cases there is at least an ultimate justification in terms of trying to understand biological systems. It is again assumed for the purposes of this paper that relevance to biology is intended.

4.3 Synthesis

Some 'animat' researchers explicitly claim that their systems are not simulations or models. For example MacLennan and Berghardt (1993) state:

> *"The techniques of synthetic ethology must be carefully distinguished from simulation or mathematical modeling. In a simulation, an attempt is made to imitate in a computer or other modeling system the salient aspects of a complex situation that exists, at least potentially, in the real world…In synthetic ethology, by contrast, we do not attempt to model any existing natural system. Instead we construct 'synthetic worlds' which are simple but complete and which manifest the phenomenon of interest" (p.161-2)*

But the 'phenomenon of interest' here is something that does 'at least potentially' exist in the real world - *"the goal is to use the synthetic approach to understand the natural phenomenon"* (op. cit.) – in this case the evolution of co-operative communication between agents. So how is it that these synthetic worlds 'manifest' the phenomenon of interest, if not by 'imitating the salient aspects' of it? For some researchers, the distinction sometimes seems to rest on the argument that (at least some) ALife systems are *realisations* of life and therefore not models[5]. Ray (1994) for example also contrasts "AL simulations" that represent some natural phenomenon with "AL instantiations" which are *"living forms in their own right, and are not models of any natural lifeform*". This has generated much philosophical speculation about whether these systems are really 'alive'; a debate that many recognise to be similar to the debate over 'strong' vs. 'weak'

---

[5] Although note that MacLennan (2002) says "we make no claim that the agents described in this report are alive in any literal sense", so does not seem to be explicitly appealing to this argument. Unfortunately they do not appear to give any alternative argument for making this distinction, other than that the synthetic worlds are very simplified and do not accurately represent real systems, but see section 5.2 below.

Artificial Intelligence. However, irrespective of the reality of Artificial Life, it is essential to note that ‘simulation’ and ‘realisation’ are not in fact exclusive categories. It is a mistake to argue that because a system is a simulation, it therefore cannot be an instantiation (Webb, 1991); and equally mistaken to argue that because a system is an instantiation, it is therefore not a model. A straightforward example is provided by the cricket robot – it does real phonotaxis, but it is also a simulation of phonotaxis in crickets.

4.4 Analogy

Langton's (1989) view of ALife as an ‘alternative’ biology provides a further twist. Clearly he believes the artificial systems to be relevant to biology, but at the same time he does not regard them as straightforward simulations of existing biological systems, or implementations of specific biological hypotheses. Rather, he appears to be arguing for them to be used as source models, i.e., analogical systems that will help in constructing novel hypotheses for biology. This is intended to proceed through a comparative approach, between the artificial and natural examples of life, to discover general principles beyond the contingent details of existing biology. Similarly, Steels (2001) suggests:

> *“we may want to compare the behaviour and mechanisms of artificial systems to that of natural systems achieving the same functionality. Comparing is not the same as mimicking or modelling the natural system ... it is a way to gain insight by confronting it with something that is different but still sufficiently similar to make the comparison interesting.” (p.1078)*

What normally characterises a ‘source’ model (or the comparative approach to constructing hypotheses) is that system being compared to the target exists, or was built, independently of the hypothesis; for example, pumps existed before they were proposed as a good model for the functioning of the heart. For some areas of ALife this might be the case, e.g., cellular automata were developed and have been explored independently of biology, but have become a popular analogy for describing how interesting global patterns might arise from local rules in biological systems, e.g. as discussed at length by Dennett (1998). Similarly, an artificial system built for an engineering purpose, or as a ‘pure’ mathematical exploration, might subsequently be usefully compared to a biological system that carries out a similar function. However, in the general area of animat models, and specifically in work such as that of Beer, the systems are not usually constructed independently, but are constructed to represent (however loosely or abstractly) some mechanisms taken, by hypothesis, to have causal relevance to biology. Consequently it seems more appropriate to consider them – to the extent that they are claimed to be relevant – as models that implement a hypothesis rather than as source models. However, either way, establishing the relevance requires some explicit specification of how the two systems – artificial and biological – are supposed to correspond.

5.5 Models

Beer (2003a) does call his simulation a model although he is somewhat equivocal regarding the intended relevance. For example he says *“The intention here is not to propose a serious model of categorical perception”* (p. 210) and yet *“The analysis described in this paper also has important implications for our understanding of perception”* (p.236), i.e., empirical claims *are* being made about the world on the basis of the model results. Arguing that *“the intent of idealized models is not empirical prediction, but conceptual clarification”* (p.240) and that *“many of the details of an idealized model’s analysis will be tied to the particular decisions made in its formulation, and thus be of questionable direct relevance to any real cognitive system”* (p.241), it seems he regards the model more in the mode of ‘pure’ exploration or potential ‘source’ analogies. Yet his system is not theory-independent, but has been specifically designed to support the assertion that *“dynamical analyses of cognitive behavior…can…broaden and clarify our understanding of the mechanisms of cognition”* (p.241). To draw any such conclusion *relies* on the fact that the agent and task have been intentionally built to represent real cognitive agents and tasks. Beer recognises this in arguing for the biological and cognitive relevance of his agent’s design and behaviour, for example when he defends the choice of CTRNNs as follows:

> *“from a scientific (as opposed to an engineering) point of view, we are not free to choose models based solely on the convenience of analysis. Rather, we are faced with understanding the particular brain-body-environment systems that Nature has presented to us” (2003b, p.302).*

That is, the task, environment and structure of the invented creature are taken to be representative of some class of real biological problem, and thus to allow the exploration of hypotheses relevant to biology. In other words, it is a biological model in the same ontological sense as the robot cricket.

4. Evaluating animat research

If it is accepted, following the previous section, that Beer’s agent – and comparable examples of animats - *are* models, then it seems reasonable ask how their behaviour and the underlying mechanisms compare to that of the target phenomenon in animals. This is surely the critical step that allows the model results to inform biology. But if the ‘target’ is a Martian three-wheeled iguana, how are we supposed to proceed? It should be evident from the previous discussion that despite the ‘invented’ nature of animats, researchers usually have in mind some real phenomenon that they wish to address. An example used by Beer (2003a) is human categorical perception, in particular, the sharp perceptual boundary between phonemes that differ on a continuous dimension

(Ehret, 1987). Indeed, he goes to the effort of showing that his agent produces a sigmoidal curve for 'labelling' (defined in terms of mean catch performance as the shape is varied systematically from a circle to a square by change in a parameter α from 0-1, fig 6A) and bell shaped curve for 'discrimination' (defined as the mean difference in catch performance for differences of $\alpha=0.1$, fig 6B), which are characteristic of categorical perception in psychophysical studies (Studdert-Kennedy *et al.*, 1970). This demonstration is rather weakened by the fact that Beer's second curve cannot possibly be anything but the differentiation of the first, whereas the critical point made by Studdert-Kennedy *et al.* is that 'categorical perception' should be distinguished from simple classification of a continuous percept by the *empirical* discovery that labelling and discrimination curves (measured using different behavioural paradigms) turn out to be so related.

But is this a fair criticism? Is it taking the model too 'seriously' to expect anomalies such as this to the natural phenomenon to be discussed? Is there something about animat research that justifies the widely encountered ambivalence towards making such direct comparisons? Again, a number of arguments can be distinguished.

4.1 Animats as existence proofs

Harvey *et al.* (2005), describing the animat approach in evolutionary robotics, say "it is not necessary … to aim at modelling specific animal behaviour so as to validate the model with empirical data" (p. 84). This is justified by the suggestion that animat simulations are meant to provide existence proofs, or proofs of concept, rather than account for data. Harvey *et al.* argue that:

> *"We will have demonstrated cognitive capacity or phenomenon X under clearly specified conditions and constraints, hence showing that these provide sufficient conditions for X… the production of proofs of concept has a different scientific function [from empirical validation]: that of catalysing theoretical re-conceptualizations and facilitating the production of novel hypotheses" (p. 84).*

A researcher could thus, for example, refute a claim that a (biological) phenomenon X requires condition Y by showing that an animat can produce X without Y.

However, 'existence proofs' clearly do require comparisons between model results and empirical data. One cannot evaluate the claim that phenomenon X requires condition Y unless one can show that phenomenon X actually is produced (with or without Y). And the claim or proof will be stronger or weaker depending on how well the simulated X matches the real X; for example, demonstrating successful behaviour in the same physical situation as the animal. Harvey et al (2005) indeed recognise that for 'proofs of concept' to be relevant *does* require an attempt at validation with empirical data – the "novel hypotheses…then need to be appropriately translated to domain-specific cases and tested empirically" (p. 84). Yet the animat conception of an invented animal often seems like a convenient way to put off this testing indefinitely. By contrast, specifying a specific animal target from the start means the model will already have one domain specific translation, and the corresponding data set, immediately to hand. Note that this does not mean a novel hypothesis produced in simulating a particular target animal system cannot subsequently be translated into other domains, and thus also evaluated for its wider application. The Hodgkin-Huxley model of neural dynamics, first developed to account for specific data from the squid giant axon, but subsequently shown to describe a very broad range of neuron types, is a notable example.

Moreover, for an existence proof to be interesting, there needs to have been some claim or general belief that phenomenon X was not possible without condition Y, for example that cricket song preference behaviour required separate sound recognition and sound localisation circuitry (Huber and Thorson, 1985) which the cricket robot demonstrated was not the case (Webb, 1995). But we did not lack for existence proofs that categorisation behaviour such as that of Beer's agent can be achieved using a simple neural network in an appropriately embodied agent interacting with the environment, without explicit representation. There were already a number of examples of simple robot systems, capable of approaching certain visual targets while avoiding other stimuli, some based specifically on the 'existence proofs' provided by insects, e.g. Srinivasan and Venkatesh, 1997. Indeed, the cricket robot itself has been used by several philosophers to illustrate exactly this point about the capabilities of such agents (e.g. Clark, 2001). And those who argue that these kinds of systems *are* in fact using representations (e.g. Bechtel, 1998; Mandik, 2002) do not find Beer's agent a convincing counterexample; the disagreement is not over how the task is actually accomplished in these simple systems but over the definition and application of the term 'representation' in describing such mechanisms.

To be fair, rather than claiming his agent is a *novel* existence proof for cognition without representation, Beer emphasises rather how his analysis is an existence proof of how a dynamical account of cognition can be made. The development and application of dynamical tools is indeed an interesting contribution, but the question of its relevance to biology remains dependent on the existence and extent of the mapping between his agent and real cognitive systems. Beer argues that he uses a minimalist invented animat because such a dynamical analysis is not yet feasible to apply to any real system, which would be far too complex – *"Only in idealized model agents can we presently hope to study the dynamics of a complete agent-environment system and thus clarify the fundamental nature of such systems."* (p. 240). But why could this not be attempted using a highly simplified or minimalist model of a real animal, rather than an invented one? The assumption seems to be that choosing a real target immediately implies one must build a complex, detailed, accurate, low level model; and show the model can reproduce all the "cognitive eccentricities" of the target; but this simply does not follow.

4.2 Animats as idealisations

Beer (2003) makes a direct connection between his work and idealised models such as the 'frictionless planes' introduced by Galileo in developing theoretical mechanics. The nature of 'Galilean idealisation' has been discussed in detail by McMullin (1985) and has as its basis the use of a simpler analogue of the original complex problem as a way of facilitating arrival at a solution. The simplification might involve 'reasonable' approximations (e.g. that the earth is flat within the distance that a projectile travels); this can be justified by arguing the results on the model are negligible. It may involve neglecting factors that are known to have a non-negligible effect on the real system (e.g. friction); this can be justified by arguing that we can estimate how this limits the conclusions of our model, and can (at least in principle) add back the necessary corrections when applying the model to a particular real situation. It may involve considering only one cause at a time in a complex multi-causal situation (e.g. the interaction of gravity and air resistance on falling bodies) and perhaps doing this in 'thought experiments' when actual experimental isolation of causes is not possible with the available technology. As Galileo said, we may find that we can "draw true conclusions even from false assumptions" (cited in McMullin, p.255).

'Galilean idealisation' is sometimes described as 'pragmatic' idealisation, i.e., the simplification has been introduced only because it is otherwise too difficult to solve the problem. This is contrasted to 'minimalist' idealisation, in which the fact that the model is a highly reduced description of the real system is taken to be a virtue in itself, allowing better understanding of the real essence of the problem (Weisberg, 2007). The term 'idealisation' seems to incorporate at least two dimensions of modelling: abstraction and inaccuracy. Perfect accuracy and no abstraction would imply a one-to-one mapping or isomorphism between the elements, structure and behaviour of the model and those of the target system[6]. Few or no actual scientific models are strict isomorphisms; the mapping is usually partial (i.e., not all elements in the system are mapped to elements of the model) and many-to-one (e.g. several steps of a process are mapped to one process in the model); these are processes of abstraction. Furthermore, in most or all models there are ways in which the mapping simply fails, rendering the model inaccurate. This may take the form of approximation, using a value that is known to be wrong but considered 'close enough', or pushing a value to a limit known to be impossible in reality, such as zero friction. It may involve substitution of one process with another that has an input-output relation that is 'similar enough' e.g. using a linear function for a monotonic

---

[6] Note this mapping could be defined at different levels (e.g. a mapping of control laws, or of brain structures, or of individual neurons). I consider the level of a model to be independent of its accuracy or abstraction.

non-linear relation. It may be a case of having elements in the model that are not intended to map to anything in the system, but need to be there to make it run. It may also involve an intentional decision to contradict some known facts about the system, either for the sake of creating a more tractable model, or indeed to directly test, in a counterfactual model, whether those particular facts are critical or not. ‘Idealisation’, then, could be thought of as inaccuracy and abstraction that limits the strict ‘completeness’ of a model in a good cause[7], e.g., to enable insights, manipulations and analyses that would not otherwise be possible.

I do not in the least wish to dispute the value of idealised models, and my critique of the animat approach is *not* that the simulations are over-simplified or falsify certain aspects of real biology. Such abstraction and inaccuracy is ubiquitous in science. For example, in running an experiment on real crickets, the environment is usually highly simplified, having flat terrain, no visual stimuli, only one or two auditory cues, and the cues themselves having much of the natural variation removed. The crickets themselves are usually pre-selected to minimise variation, using one species, at a certain age, in good condition etc., and the reported behaviours are combined across individuals and trials. Thus even the data that we try to account for is already simplified and distorted model of the ‘normal’ behavioural situation and response. The cricket robot then further simplifies and distorts in its representation of real crickets. This is perhaps most obvious in the motor output, where two wheels substitute for six legs. It also has one auditory transducer on each side, compared to the cricket’s 50 or so receptor cells, and these transducers differ in a number of other respects. The neural circuit does not include a number of identified auditory neurons that may influence phonotaxis (such as the thoracic neuron pair labelled ‘AN2’, Schildberger and Horner, 1988). The output of the robot provides only a qualitative match to real cricket data. For example, the discrimination or preference curve for songs of different syllable rates for the robot has the same characteristic band-pass shape, but it is not value by value identical to the curve for the cricket.

It is also well established in the philosophy of science that a great deal can be learnt from models that are known to misrepresent the target system (arguably, this includes all models in science – Cartwright, 1983). Wimsatt (1987) discusses the issue in terms of “False models as a means to truer theories”, listing twelve uses for ‘false’ models that include: acting as a starting point for developing better models; undercutting a preferred hypothesis by indicating an alternative line of explanation; as a neutral baseline for assessing causal claims in other models; as a phenomenological fitting of data that might suggest an underlying mechanism, or indicate which factors have most predictive power; by comparing which results hold across a family of models that

---

[7] Of course, models are often inaccurate or abstract for more mundane reasons, such as lack of facilities or knowledge, without thereby being considered ‘ideal’ for any purpose.

have differing false assumptions. Frigg (2003) adds several points, such as how we may learn during the building process, forcing us to see the target system from a new point of view; and that features may be highlighted by isolation or exaggeration. Steels (2001) in discussing ‘formal’ (idealised?) models approvingly quotes the economic modeller Milton Friedman: “*to be important. ... a hypothesis must be descriptively false in its assumptions*”. Two uses for 'idealised models' that are often stressed in AB and ALife are that we can search for the minimum set of components that still produces something like the phenomenon of interest, and that this might be found by some automatic search process, typically a genetic algorithm, to generate solutions that are relatively unbiased by the modeller's theoretical preconceptions (Bedau, 1999). Harvey *et al.* (2005) note that this work might be independent of any claim that “this gives us direct insights into the actual physical mechanisms of real organisms” (p.84).

But if the results of animat models are to be relevant to biology, then ultimately *some* claim about the physical mechanisms of real animals must be asserted. There are at least two ways in which the animat approach limits such claims. The first is that inventing or imagining a system is not necessarily the same as devising an idealised representation of a real system. There is no basis for assessing what has been abstracted or distorted, or whether assumptions in the hypothesis are false or true, if there is no identified target to which the model can be compared. One cannot provide a coherent argument for why the factors that have been ignored are negligible, or a systematic outline of how non-negligible factors could be gradually reintroduced to allow the real situation to be more closely represented, or what facts the model is counterfactual to. The second is that lacking an *a priori* target, the animat researcher can choose *post hoc* to note any biological data that seems to match their results while conveniently ignoring any data that does not.

4.3 Animats as generalisations

One possible response to the above criticism regarding the lack of clear targets in animat models is that they do indeed have real targets, but these are not limited to some specific organism; rather, what is represented by the model is some general problem of biology or cognition. Moreover, in contrast to tightly targeted modelling such as the cricket robot, this very lack of specificity means that the model is more likely to provide some fundamental conceptual insights. Indeed, this idea is explicit in Dennett’s original (1978) suggestion to “*make up* a whole cognitive creature” in order “to uncover *very* general, *very* abstract principles”.

Generality is sometimes contrasted to being definite or precise, for example when we talk about something in ‘general terms’, meaning we are using rough approximations. But this meaning of general does not seem to apply to animat models which usually *are* precisely defined, with the exact properties of inputs and outputs and the equations of transformations fully described, as they need to be for the simulated or physically implemented system to be run. On the other hand, ‘generality’ can be contrasted to specificity; when we talk about the properties common to a range of systems, rather than those specific to only one system. Unlike the previous definition this explicitly invokes the idea of a description that covers multiple systems; you cannot talk about a ‘general’ description of only one system. This idea of generality is more obviously the one that is valued in science. To be able to represent multiple systems with one model implies unification of explanation, a widely accepted criteria for a good scientific theory.

How can this desirable generality be obtained? We should first dispose of the common but mistaken idea that abstraction *directly* implies generality, with all its virtues: e.g. Bedau (1999) in justifying simple ALife models says “the more abstract the model, the more broadly it applies, so the more unified the explanation it provides” (p.20). It is true that the subset of properties general or common to a number of different systems must necessarily be no larger than the set of properties of any one of those systems, and thus a general description is likely to contain fewer elements, i.e., be more abstract. But the opposite is not automatically true: an abstract model, containing few details, is not necessarily general. It may still only describe one system, or indeed no existing system at all. Similarly it is important not to assume that generality must be associated with higher levels of description. Lower level principles in biology are more general: biochemistry encompasses a larger range of systems than neuroscience; and neuroscience a larger range than cognitive science. Lower level models often *are* elaborations of higher level ones and hence less abstract. But this does not necessarily make them less general, as these lower level details may still be common to all the systems being targeted – e.g. the neural level cricket model applies to just as many cricket species as the algorithmic version did. And lower level models are not always less abstract. Neural network models of perceptual classification were interesting precisely because they seemed much simpler than the higher-level rule-based models that had been previously proposed: the lower level was *not* just implementation detail but an entirely different, simpler, and more abstract, principle of operation. Much of the interest in ALife and AB comes from just this kind of explanation, in which some abstract low-level principle accounts for high-level complexity.

However, a more interesting association of ‘high-level’ and ‘generality’ is the idea that we can find general principles of operation turning up at higher levels of description *despite* the intervening levels having different principles of operation. We assume both computers and animals are governed at the lowest level under the same general laws of physics; general principles at the level of biochemistry are not relevant to computers; yet we assume the principles governing networks of neuron-like units are the same for computer simulations and for real neurons. For example, we assume that there is nothing at the level of biochemistry that is essential to performing ‘real’ phonotaxis, thus a robot can not only represent but reproduce the relevant mechanisms in the cricket. This idea, of higher level capabilities that can be multiply realised in systems that differ in their lower level properties, is of course central to ALife (and to Artificial Intelligence and Cybernetics before it):

> *"Of course, the principle assumption made in Artificial Life is that the 'logical form' of an organism can be separated from its material basis of construction, and that 'aliveness' will be found to be a property of the former, not of the latter." [Langton, 1989, page 11].*

It also supplies a potential motivation for the animat approach. As spelt out by Langton, the idea is that by inventing or artificially constructing novel examples of the phenomenon of interest ('life' or 'adaptive behaviour') we discover the general (high-level) laws of life-as-it-could-be rather than the contingent (low-level) specifics of life-as-it-is. The problem remains, however, of how to decide that the invented system *is* an example of the same phenomenon, if we lack these general laws. This issue does not seem too problematic for phonotaxis, but is hotly debated for 'life' and 'cognition'. Moreover, as Keeley (2000) usefully points out, there seems to be an inherent circularity in which we decide the taxonomy of a phenomenon without reference to lower levels, and then take this as proof that the lower levels are irrelevant to the phenomenon. It remains a possibility that to replicate life – or cognition – requires the right artificial chemistry, for example. Discovering principles relevant to all life is a laudable aim, but these may turn out to be contingent after all. It should be noted, then, that it *is* possible to aim for general explanations (e.g. of all existing examples of life or cognition) without assuming multirealisability (i.e., that life could also occur with some different material basis).

Again, it is not the potential value of general models that I wish to dispute. Rather, I wish to question the basis for the assertion that animat models are general. An animat is a made-up creature (usually with its own unique set of "cognitive eccentricities", such as having exactly seven distance sensors, and only being able to move horizontally) that has been specifically designed not to represent any real animal. Why should we then suppose that it somehow, at the same time, represents more biological systems than a simulation that is explicitly based on at least one existing species? To show that a 'general' model targets many systems still requires that these systems be specified, that the mapping (however partial and inaccurate) between the model and the real systems be described, and the behaviour compared. Evolution itself is a good example of a general hypothesis, asserted to apply to all biology, but supported by the exhaustive demonstration of its applicability in explaining a huge array of specific cases.

In fact, Beer does not explicitly claim his 'idealised' model is a general one in the sense discussed so far. Rather, he says it is put forward as a specific example of a 'minimally cognitive system', and that close examination of this specific system will be a route by which general principles about cognition may emerge. Just the same motivation is given (e.g. Reeve and Webb, 2003) for the cricket robot: phonotaxis is being studied in detail not simply to understand more about crickets, but because it represents many general problems of sensorimotor control. Indeed, these are largely the same problems that interest Beer – such as the interaction of body, task and environment, the role of individual neurons in behaviour, and what minimal complexity of behaviour requires cognitive capabilities.

5 Conclusion

I have argued that animat simulations, if they are to be relevant to biology, should be considered as models. That they are abstract, approximate and not necessarily accurate; that they are used mostly for exploring general principles, providing existence proofs and generating novel hypotheses; does not remove the requirement to demonstrate that they can be usefully compared to observations made on real biological systems. The practice of inventing a non-existent animal, rather than specifying the real animal system(s) that are the target of the model, tends to undermine the fulfilment of this requirement. For example, it allows critique of any particular element of the representation to be deflected, and rigorous comparison to empirical data to be indefinitely postponed. But unless this requirement is fulfilled, work on animats can be considered only as the exploration of artificial systems – perhaps interesting in their own right, or with potential technological applications – but not part of biology.

It is noticeable that many of the arguments for the animat approach are based on 'making the best the enemy of the good'. Thus, it is often argued that because it is impossible to provide detailed, accurate and complete models of any real animal system, the only alternative is to invent artificial creatures for study. But this ignores the perfectly plausible option of building abstracted, inaccurate and incomplete models of real animals. As I have argued here, there is little difference between my cricket robot and Beer's minimally cognitive agent on most of the dimensions of modelling. The level of description is much the same, i.e., neural activity supporting discriminatory motor behaviour. The robot model abstracts and distorts many elements and processes in the real cricket, and only qualitatively matches cricket data. It is assumed by both models that general insights into sensorimotor control, relevant to biology, can be obtained from these specific examples. Although the medium (robot or computer) differs, it is not hard to imagine a purely computational version of the cricket model, or a robotic implementation of Beer's agent. The cricket robot provided a novel view of, and new hypotheses for, the perceptual task performed by the animal, and an existence proof that no explicitly representational processing is needed for this “minimally cognitive” task[8].

The only apparent advantage in not grounding one's model by reference to some real target system (i.e., for using animat rather animal models) is that the model is thus immune to criticism of its validity. The price paid for this immunity, however, is to undermine any justification for its relevance. It is often forgotten in discussions of Braitenberg's (1984) thought experiments about ‘hypothetical’ vehicles that he spends the second half of his book explaining the biological relevance of each of the features with which they have been endowed. To do so requires reference to real iguanas, not the Martian three-wheeled variety.

---

[8] In fact, I do not wish to claim the cricket task is 'cognitive', but it is no less 'cognitive' than the task performed by Beer's agent.

References

Andersson, S.B., Handzel, A.A., Shah, V., and Krishnaprasad, P.S. 2004. Robot Phonotaxis with Dynamic Sound Localization, *Proceeding IEEE International Conference on Robotics and Automation (*pp. 4833-4838)

Arbib, M. 1982. Rana Computatrix: An Evolving Model of Visuo-motor Coordination in Frog and Toad. In J.E. Hayes, D. Michie, and Y.H. Pao (eds.) *Machine Intelligence 10* (pp. 501-517) Chichester: Ellis Horwood Publishers.

Bechtel, W. 1998. Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge in Cognitive Science. *Cognitive Science,* 22:295-318.

Bedau, M. A. 1999. Can Unrealistic Computer Models Illuminate Theoretical Biology? In A. Wu, (ed.) P*roceedings of the 1999 Genetic and Evolutionary Computation Conference (*pp. 20-23). Orlando: GECCO.

Beer, R.D. 1996. Toward the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J.A. Meyer, J. Pollack and S. Wilson (eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (pp. 421-429). Cambridge MA: MIT Press.

Beer, R.D. 1997. The dynamics of adaptive behavior: A research program. *Robotics and Autonomous Systems,* 20:257-289.

Beer, R.D. 2003a. The dynamics of active categorical perception in an evolved model agent *Adaptive Behavior,* 11(4):209-243.

Beer, R.D. 2003b. Arches and stones in cognitive architecture. *Adaptive Behavior,* 11(4):299-305.

Böhm, H., Schildberger, K., Huber, F. 1991. Visual and acoustic course control in the cricket *Gryllus-bimaculatus*. *Journal of Experimental Biology,* 159:235-248.

Bongard, J. and Paul, C. 2000. Investigating Morphological Symmetry and Locomotive Efficiency using Virtual Embodied Evolution. In J.-A. Meyer et al (eds.) *From Animals to Animats: The Sixth International Conference on the Simulation of Adaptive Behavior.*

Braitenberg, V. 1984. *Vehicles: Experiments in Synthetic Psychology.* Cambridge MA*:* MIT Press.

Brooks, R. A. 1989. How To Build Complete Creatures Rather Than Isolated Cognitive Simulators. In K. VanLehn (ed.) *Architectures for Intelligence* (pp. 225–239), Erlbaum: Hillsdale, NJ.

Brooks, R. A., 1990. Elephants Don't Play Chess. *Robotics and Autonomous Systems,* 6:3–15.

Brooks, R. A., and Stein, L. A. 1994. Building Brains for Bodies, *Autonomous Robotics,* 1 (1):7-25

Calenbuhr, V. & Deneubourg, J.L. 1990. A model for trail following in ants: Individual and collective behaviour. In W. Alt and G. Hoffman (eds.) *Biological Motion*, Lecture Notes in Biomathematics, (pp.453-469) Berlin: Springer-Verlag.

Callender, C. and Cohen, J. 2005. There Is No Special Problem About Scientific Representation. http://philsci-archive.pitt.edu/archive/00002177/

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

Collins, S., Ruina, A., Tedrake, R., Wisse, M. 2005. Efficient Bipedal Robots Based on Passive Dynamic Walkers. *Science Magazine,* 307:1082-1085.

Clancy, W. J. 1997. *Situated cognition*. Cambridge: Cambridge University Press.

Clark, A. 1997. *Being There: Putting Brain, Body, and World Together.* Cambridge MA: MIT Press.

Clark, A. 2001. *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: OUP.

Dennett, D.C. 1978. Why not the whole iguana? *Behavioral and Brain Sciences,* 1: 103—104.

Dennett, D.C.1998. *Brainchildren: Essays on Designing Minds.* Cambridge, MA: MIT Press

Dorigo, M. and Stützle, T. 2004. *Ant colony optimization*. Cambridge, MA: MIT Press.

Ehret, G. 1987. Categorical Perception of Speech Signals: Facts and Hypotheses from Animal Studies. In S. Harnad (ed.) *Categorical perception: The groundwork of Cognition.* New York: Cambridge University Press.

Frigg, R. 2003. Self organised criticality—what it is and what it isn't. *Studies in the History and Philosophy of Science* 34:613–632.

Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldenberg, E., Garcowski, J. and Beer, R.D. 2004. May We Have Your Attention: Analysis of a Selective Attention Task. In *From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior,* July 13-17, 2004, Los Angeles, CA.

Guillot A. and Meyer, J-A. 1997. Synthetic animals in synthetic worlds. In Kunii et Luciani (Eds). *Synthetic Worlds*. John Wiley and Sons.

Harvey, I., Di Paolo, E. A., Tuci, E., Wood, R., Quinn, M., 2005. Evolutionary robotics: A new scientific tool for studying cognition. *Artificial Life,* 11:79 - 98.

Hedwig B., Poulet J.F.P. 2004. Complex auditory behaviour emerges from simple reactive steering. *Nature,* 430:781-785.

Hedwig B., Poulet J.F.P. 2005. Mechanisms Underlying Phonotactic Steering in the Cricket *Gryllus bimaculatus (de Geer)* Revealed with a Fast Trackball System. *Journal of Experimental Biology*, 208: 915-927.

Hennig R.M., Franz A., Stumpner A. 2004. Auditory processing in insects. Microscopy Research Techniques, 63:351 – 374.

Horchler, A., Reeve, R., Webb, B. & Quinn, R. 2004. Robot phonotaxis in the wild: A biologically inspired approach to outdoor sound localization. *Advanced Robotics,* 18(8):801-816.

Huber, F. and Thorson, J., 1985. Cricket Auditory Communication. *Scientific American*, 253:47-54.

Husbands, P., Harvey, I. and Cliff, D. 1995. Circle in the Round: State Space Attractors for Evolved Sighted Robots. *Robotics and Autonomous Systems* 15:83—106.

Keeley B. L. 2000. Shocking lessons from electric fish: The theory and practice of multiple realization. *Philosophy of Science,* 67:444-465 .

Kim, K-J. and Cho, S-B. 2006. A comprehensive overview of the applications of ALife. *Artificial Life,* 12:153-182.

Koch, C. 1999. *Biophysics of Computation*. Oxford: Oxford University Press.

Langton, C. 1989. Artificial Life. In C. Langton (ed.) *Artificial Life* (pp. 1-47). Addison-Wesley.

Lund, H. H., Webb, B., & Hallam, J. 1997. A robot attracted to the cricket species *Gryllus Bimaculatus*. In *Fourth European Conference on Artificial Life* Cambridge MA: MIT Press

Lund, H. H., Webb, B., & Hallam, J. 1998. Physical and temporal scaling considerations in a robot model of cricket calling song preference. *Artificial Life,* 4(1):95-107.

MacLennan, B. 1992. Synthetic Ethology: An Approach to the Study of Communication. In C. G. Langton, C. Taylor, J.D. Farmer and S. Rasmussen (eds.) *Artificial Life II: The Second Workshop on the Synthesis and Simulation of Living Systems* (pp. 631-658) Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Vol. X. Redwood City, CA: Addison-Wesley.

MacLennan, B. and Burghardt, G. M. 1993. Synthetic Ethology and the Evolution of Cooperative Communication. *Adaptive Behavior*, 2(2):161-188.

MacLennan, B. 2002. Synthetic Ethology: A New Tool for Investigating Animal Cognition. In Marc Bekoff, Colin Allen, and Gordon M. Burghardt (eds.), *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition* (pp. 151-156). Cambridge MA: MIT Press.

McMullin, E. 1985. Galilean idealization. *Studies in History and Philosophy of Science,* 16:247-273

Mandik, P. 2002. Synthetic Neuroethology. *Metaphilosophy*. 33 (1-2):11-29.

Meyer J.A., Guillot A. 1991. Simulation of Adaptive Behavior in animats: Review and Prospect. In Meyer J.A., Wilson S.W. (eds.) *From animals to animats. Proceedings of the First Conference on Simulation of Adaptive Behavior (pp.2-14)*. Cambridge MA: MIT Press.

Meyer, J.A., Guillot, A., Girard, B., Khamassi, M., Pirim, P. and Berthoz, A. 2005. The Psikharpax project: Towards building an artificial rat. *Robotics and Autonomous Systems,* 50(4):211-223

Michelsen, A., Popov, A.V., Lewis, B. 1994. Physics of directional hearing in the cricket *Gryllus bimaculatus*. *Journal of Comparative Physiology A*, 175:153-164.

Neumann, T.R. and Bülthoff, H.H. 2002. Behavior-oriented vision for biomimetic flight control. P*roceedings of the EPSRC/BBSRC International Workshop on Biologically Inspired Robotics: The Legacy of W. Grey Walter* (pp.196-203), 14-16 August, HP Labs Bristol, UK.

Neumann, T. R., Huber, S. and Bülthoff, H. H. 2001. Artificial systems as models in biological cybernetics. *Behavioural and Brain Sciences,* 24 (6):1071-1072.

Pfeifer, R. and Scheier, R. 2000. *Understanding Intelligence.* Cambridge MA: MIT Press.

Poulet J.F.A., Hedwig, B. 2005. Auditory orientation in crickets: Pattern recognition controls reactive steering. *Proceedings of the National Academy of Science,* 102:15665-15669.

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

Reeve, R. & Webb, B. 2003. New neural circuits for robot phonotaxis. P*hilosophical Transactions of the Royal Society A,* 361:2245-2266.

Reeve, R., Webb, B., Indiveri, G., Horchler, A., & Quinn, R. 2005. New technologies for testing a model of cricket phonotaxis on an outdoor robot. *Robotics and Autonomous Systems,* 51(1):41-54.

Schildberger, K. 1984. Temporal selectivity of identified auditory interneurons in the cricket brain. *Journal of Comparative Physiology A*, 155:171-185.

Schildberger, K, Horner, M. 1988. The function of auditory neurons in cricket phonotaxis .1. Influence of hyperpolarization of identified neurons on sound localization. *Journal of Comparative Physiology A*, 163, 621-631.

Scholes, S., Wilson, M., Sendova-Franks, A.B. and Melhuish, C. 2004. Comparisons in Evolution and Engineering: The collective Intelligence of Sorting. *Adaptive Behaviour,* 12: 147-159.

Simon, H. A. 1969. *The Sciences of the Artificial*. Cambridge MA: MIT Press.

Slocum, A.C., Downey, D.C. and Beer, R.D. 2000. Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J. Meyer,

A. Berthoz, D. Floreano, H. Roitblat and S. Wilson (eds.), F*rom Animals to Animats6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior* (pp. 430-439), Cambridge MA: MIT Press.

Srinivasan, M. V. & Venkatesh, S. (1997). *From living eyes to seeing machines*. Oxford: Oxford University Press.

Steels, L. 2003. Evolving Grounded Communication for Robots. *Trends in Cognitive Sciences*, 7(7):308--312.

Steels, L. 2001. **The methodology of the artificial.** *Behavioral and Brain Sciences*, 24:1077-1078.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S. and Cooper, F. S. 1970. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review* 77:234–249.

Taylor, P. 1989. Revising models and generating theory. *Oikos*, 54:121-126.

Voegtlin, T. and Verschure, P.F.M.J. 1999. What can robots tell us about brains? A synthetic approach towards the study of learning and problem solving. *Reviews in the Neurosciences* 10(3-4):291-310.

Verschure, P.F.M.J., Voegtlin, T. & Douglas, R.J. 2003. Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature*, 425: 620-624.

Walter, W. G. 1950. An imitation of life. *Scientific American* 182(5): 42—45

Walter, W. G. 1951. A machine that learns. *Scientific American* 185(2): 60—63

Webb, B. 1991. Do computer simulations really cognize? *Journal of Experimental and Theoretical Artificial Intelligence,* 3:247-254.

Webb, B. 1995. Using robots to model animals: a cricket test . *Robotics and Autonomous Systems*, 16, 117-134.

Webb, B. 2000. What does robotics offer animal behaviour? *Animal Behaviour* 60:545-558.

Webb, B. 2001. Can robots make good models of biological behaviour? *Behavioural and Brain Sciences* 24 (6):1033-1050.

Webb, B. 2006. Validating biorobotic models. *Journal of Neural Engineering* 3 R25-R35

Webb, B., Harrison, R. R., 2000. Integrating sensorimotor systems in a robot model of cricket behavior. *Sensor Fusion and Decentralised Control in Robotic Systems III* . Boston Nov 6-8: SPIE.

Webb, B., & Scutt, T. 2000. A simple latency dependent spiking neuron model of cricket phonotaxis. *Biological Cybernetics,* 82(3):247-269.

Weisberg, M. 2007. Three kinds of idealization. *Journal of Philosophy*, CIV(12): 639-659.

Wilson, S.W. 1985. Knowledge growth in an artificial animal. *Proceedings of the First International Conference on Genetic Algorithms and Their Applications* (pp. 16-23). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Wilson, S.W. 1991. The Animat Path to AI. In J. A. Meyer and S. W. Wilson (eds.), *From Animals to Animats: Proceedings of the First InternationalConference on the Simulation of Adaptive Behavior*, Cambridge, Massachusetts: The MIT Press/Bradford Books

Wimsatt, W.C. 1987. False Models as means to Truer Theories. In: M. Nitecki and A. Hoffman, (eds.). *Neutral Models in Biology*. London: Oxford University Press

Wohlers D.W., & Huber F. 1982. Processing of sound signals by six types of neurons in the prothoracic ganglion of the cricket, *Gryllus campestris L. Journal of Comparative Physiology A*, 146: 161–173.