Brit. J. Phil. Sci. **48** (1997), 587–601

REVIEW ARTICLE

Universal Biology[1]

Kim Sterelny

1 Introduction: the N = 1 problem

Like many collections, these two new surveys of A-Life are variable in theme and quality. Both contain very interesting papers, but also others that are dull. The two collections have rather different aims. Langton's is mostly a collection of overviews of various subdomains of A-Life, reviewing achievements and outstanding open problems, and speculating on future developments. As one might expect, Boden's is more focused on philosophical implications, and is richer in extended critical discussion. By and large, only papers in the Boden collection argue rather than report. In what follows I will extract and discuss what seem to me to be some main themes common to the two collections.

On even a cursory survey of A-Life literature, one is struck by the resuscitation of a quaintly old-fashioned project: defining life. Definitions have gone out of fashion lately. In semantics and psychology, it is now realized that our capacity to use concepts and refer to kinds does not depend on a grasp, implicit or explicit, on the necessary and sufficient conditions of membership of that category. Though natural kinds may have essences, they are discovered not through the construction of definitions at the beginning of inquiry, but, if we are lucky, as their culmination. So why suppose that biology needs a definition of life? Having a definition will not help with odd and hard-to-classify cases like viruses and social insect colonies, for the adequacy of the definition is settled by our view of the cases, not vice versa.

I do not think players in the A-Life field take definitions of life merely to be props helping us understand what we are talking about. Consider, for example, Bedau's paper (in Boden). He begins with Mayr's characterization of life:

[1] Review of C. Langton (*ed.*) [1995]: *Artificial Life: An Overview*, MIT Press, and M. Boden (*ed.*) [1996]: *The Philosophy of Artificial Life*, Oxford University Press.

> 1. All levels of living systems have an enormously complex and adaptive organisation.
>
> 2. Living organisms are composed of a chemically unique set of macromolecules.
>
> 3. The important phenomena in living systems are predominantly qualitative, not quantitative.
>
> 4. All levels of living systems consist of highly variable groups of unique individuals.
>
> 5. All organisms possess historically evolved genetic programs which enable them to engage in 'teleonomic' processes and activities.
>
> 6. Classes of living organisms are defined by historical connections of common descent.
>
> 7. Organisms are the product of natural selection.
>
> 8. Biological processes are especially unpredictable (Boden, p. 336).

As we all would, Bedau accepts this as a good initial characterization of life. *But only of life as we find it.* He wants a guarantee that the traits on the list are on it by necessity not by chance. That guarantee can only come, if anywhere, from finding a master variable that explains *why* life has these characteristics. Bedau himself has a candidate to run: 'supple adaptation'. For Bedau, lineages that evolve under natural selection are the primary living things. That master variable can then serve as a theoretical rather than common sense definition of life. Bedau's approach is characteristic of the field. Many of its practitioners have an overriding interest in 'universal biology'; the biology of any life anywhere. Moreover, they argue that the development of universal biology is profoundly impaired by the 'N = 1 problem'.

> The set of biological entities provided to us by nature, broad and diverse as it is, is dominated by accident and historical contingency. We trust implicitly that there were lawful regularities at work in the determination of this set, but it is unlikely that we will discover many of these regularities by restricting ourselves only to the set of biological entities that nature actually provided. Rather, the regularities will be found only be exploring the much larger set of possible biological entities (Langton, p. x).

and

> Ideally, the science of biology should embrace all forms of life. However, in practice, it has been restricted to the study of a single instance of life, life on earth. Because biology is based on a sample size of one, we cannot know what features of life are peculiar to earth, and what features are general, characteristic of all life (Boden, p. 111).

One point of A-Life is to increase N and, in doing so, generate a 'definition of life'; that is, to tell us which features of life are essential to life in and of itself. There is one uncontroversial way of understanding this project. Suppose we could construct functioning 'organisms' with DNA, but where the DNA–RNA–protein coding relations were different. If our 'plants' grew and photosynthesized, and our 'animals' moved, ate the 'plants' and one another, and if these organisms breed true, we would thereby learn that the (almost) universal DNA–protein coding relation is not necessary. Whether it is 'chance' is another matter. Similarly, if we could build creature-like physical systems with a metabolism, physical boundaries, and the capacity to act, survive, and reproduce using macromolecules other than those found in natural organisms, we would know that our suite of macromolecules is not necessary. There is work of this kind, and it is reviewed—rather technically—by Schuster (in Langton). There are interesting results: an extra base pair is chemically possible, and that does raise the interesting question about why life has a 4-letter rather than 6-letter alphabet.

However, most who seek to ameliorate the $N = 1$ problem do not have in mind *in vitro* life. They have in mind increasing N on the cheap. Just as 'Strong AI' claims that some computing systems housed in current or near-current generation computers are not mere simulations of thought and thinking but exemplifications of it, they argue that some computer models of lifelike interactions are not simulations of life but instances of it. The data structures in, for example, Ray's famous 'Terra' program are alive, not merely illustrations of life (See, especially, Langton and Ray; for a critique, see Sober—all in Boden).

I have two indirect comments to make on this proposal before taking up the challenge of 'Strong A-Life'. First, I doubt that $N = 1$ is a problem now, or will be any time soon. The problem of 'universal biology' can be attacked by the construction of distinct theories which have different implications for evolutionary, developmental, and ecological possibility, and which can be tested by their different implications for the huge and varied experiment we actually have. We do not have a wonderful array of theories that are well confirmed and empirically equivalent with respect to the N we have, but with different implications about how life might have been. $N = 1$ may begin to bité if and when we have to decide between empirically well-confirmed and locally equivalent theories of life right here. But we are yet to recline on that couch of luxury.

Second, there is a very good question lurking behind the idea of a universal biology. We do want an explanation not just of actual biology in all its diversity but also an account of why that diversity is not greater still. But it seems to be a mistake to conceptualise lists like that of Mayr by contrasting chance with necessity. Raup has used simulation techniques to construct a representation of

lifelike shell shapes. To a first approximation, shell form can be represented as the outcome of only three different growth parameters. He shows that actual shells occupy a rather small region of the space of possible shells (for a very elegant discussion, see Dawkins [1996], Chapter 6). Why? Is this restriction a consequence of function, of subtle constraints on development, or of historical contingency? These are clearly difficult but interesting questions. But it is surely unlikely that most of the unoccupied region is flat impossible or that the occupied region is occupied through nothing but historical chance. So as Dennett [1995] has noted, contrasting chance with necessity is likely to be the wrong way of posing this problem. Similarly, there are no species with three sexes, and that is no accident. As the literature on the evolution of sex makes amply clear, sex has a cost, and that cost would increase with the number of sexes. But should we infer that the evolution of three sexes is impossible? That would surely be rash: we can conceive of a developmental biology that might work with three sexes. Nuclear DNA had two parents, so we could have three if mitochondrial DNA came from a third. But an evolutionary trajectory leading to triple sex would be both available to a lineage and favoured by selection only in very extraordinary circumstances.

2 The 'strong program' in A-Life

In philosophy of mind, there is a standard distinction between the 'strong' and 'weak' program in AI research. The weak program embraces computational models of psychological processes as a critically important research tool, but thinks of these models as theories of cognitive process. They are tools for studying the mind. In contrast, the 'strong program' in AI takes computational models to be instances, not merely representations, of cognition. The 'strong program' in A-Life suggests that computational models of evolution and ecology (if they meet suitable conditions) exemplify rather than merely represent life. On the face of it this idea is very implausible. Strong AI, after all, has a clear motivation. It begins from a computational theory of cognition. On that view, intelligence just is the capacity to execute computational tasks of suitable complexity and importance, and thinking just is the computational manipulation of symbolic structures (see e.g. Pylyshyn [1984]). Surely, computers really do compute rather than merely simulate computation. Why else would we call them computers? So there can be no principled reason why computers cannot *exemplify* thought. The ones we have now might not be fancy enough, or the computation they do might not be of the right sort. But if they fail to think, that is for contingent, perhaps passing, reasons.

But why might one think A-Life programs, or the denizens they create, are alive? The answer lies in a second route to Strong AI. Consider the following extract from an A-Life manifesto:

> Life is a property of form, not matter, a result of the organisation of matter rather than something that inheres in the matter itself. Neither nucleotides nor amino acids nor any other carbon-chain molecule is alive—yet put them together in the right way, and the dynamic behaviour that emerges out of their interactions is what we call life. It is effects, not things, upon which life is based—life is a kind of behaviour, not a kind of stuff—and as such it is constituted of simpler behaviours, not simpler stuff (Boden, p. 53).

These suggestions trade on a second defence of Strong AI, one routed through functionalism in the philosophy of mind. On one way of reading the last thirty years of philosophy of mind, the downfall of the Identity Theory was driven by the 'discovery' of the multiple realizability of mental kinds and mental properties. Just as there are many different physical systems that are all, none the less, mousetraps, so too there are many different physical systems that exemplify the same psychological properties. The 'essence of mind' is not being made of the right matter or materials. It is form, organization, or function: some relatively abstract property. Because the essential features of having a mental life are not tied to a specific physical implementation, psychological properties are 'substrate-neutral'. Mental properties are functional properties, not physical ones. Hence very different physical systems, and in particular computers, can have the right abstract, organizational features despite being brain-free.

Godfrey-Smith (in Boden) shows that those who have theorized about life and mind have often thought there was an essential similarity between life and mind: 'Life and Mind have a common abstract pattern or set of basic organisational properties. The functional properties characteristic of mind are an enriched version of the functional properties that are fundamental to life in general. Mind is literally life-*like*' (Boden, p. 320). Strong A-Life is another defence of this idea. Just as mental properties are 'substrate-neutral' so too (they claim) are critical biological properties. Both evolution and replication are substrate-neutral, for at base they consist in copying, spreading, changing and using information. Just as replication is the replication of information for building the vehicle of the next generation, development is the execution of the instructions of that replicated program.

Even granted the terms of the debate, it is far from obvious that these critical biological processes are substrate neutral; matters of form or information rather than stuff. For example, Susan Oyama and others who have developed 'developmental systems theory' have campaigned long and hard against the view that development is the execution of a genetic program (Oyama [1985]; Griffiths and Gray [1994]). Moreover, if the program-execution conception of development is flawed, it drags down with it the information-preserving conception of replication. For the information allegedly conserved is information about how to make a phenotype, which then recopies the genotype.

Most importantly, we should *not* grant the terms of this debate. In retrospect, it's amazing how easily this view of multiple realization and its significance slipped through to consensus within philosophy of mind. Achieving consensus depended on an artful mix of examples. One set of examples focus on different creatures who really do all exemplify the same psychological characteristics. Chimps, humans with standard brain architectures but exhibiting ordinary micro-variation in neural organization, and at least some humans with unusual brain architectures can all, for example, feel pain or fear. Most educated adults, despite microvariation throughout their central nervous system, believe the earth is a sphere. That belief is a shared mental property. In these cases, there are shared mental properties despite some physical differences, but for all we know there may be an underlying physical identity. Other examples are of wholly imaginary beings: in these we are invited to consider the possibility of minds without the right stuff—Martians, robots, and the like. But what evidential force do these purely imaginary examples have? So the multiple realizability of mental properties is a plausible conjecture that has been elevated to a fact.

Much more importantly, as Lycan has repeatedly argued, the stuff/form, matter/organization, implementation/function picture is a mistaken dichotomy (Lycan [1990]). There is no single level of functional or organizational description contrasted with a description of implementation. Rather, there is a cascade of increasingly or decreasingly abstract descriptions of the one system. Lycan defends this view in detail in philosophy of psychology. It seems to hold good in biology as well. For some purposes, a highly abstract, purely informational description of the genome may indeed be appropriate. For others, we may want to know in great physical detail the structure of the DNA molecule; when, for example, we want to explain its coiling properties. Other needs will call for intermediate degrees of detail—a fact already recognised in the discipline by a proliferation of distinct notions of 'the gene'. So the 'substrate neutrality' thesis rests on a false dichotomy: either we must conceive of replication in the most abstract informational terms or we must conceive it in its most intimate molecular detail. There is a whole language of genetics—of introns and exons, of crossing-over, of gene duplication and gene repair—that is functional in abstracting away from the intricate details of molecular mechanism—some of which are still, indeed, not known—yet is not purely informational.

If evaluating 'the strong program' were all there were to the philosophy of A-Life, it would be stupendously dull. Fortunately, A-Life has, potentially, a real role to play in trying to understand how entrenched features of life are. But in thinking so we only have to treat A-life models as representations of life. In running simulations, we are trying to find what those models/theories predict, when those predictions are inaccessible to analytic techniques (see Taylor *et al.*

in Langton). The great virtue of these simulations is that one can play with various parameters and thus get a feel for what outcomes are robust under fine-scale changes in the model, and which are not. Thus, for example, Nilsson's model of the evolution of eyes is impressive because the parameters are chosen conservatively, and yet eyes evolve, by geological standards, with great speed (Nilsson and Pelger [1994]; see also Dawkins [1996], Chapter 5). Simulations are important, but extracting their message does not require us to think of them as actually alive.

3 Simulation as representation

A-Life simulations are then representations of biological processes. As such, what has been their distinctive role? I take interesting A-Life simulations to fall into three classes. The first is that of simulations of relatively well-understood biological phenomena; for example, of simple predator/prey or host/parasite interactions, or even of spider webs and flies.[2] These are useful not for what they tell us of life but for what they tell us of the models. They serve to calibrate the models. If simulations of host/parasite interaction did not show the evolution of host resistance to parasites, and parasite response to changing host defences, we would not conclude that immune systems and the like are an accidental quirk of life on earth. We would rightly throw away the models as useless.

A second use of A-Life models would be as disciplined probes of the possible. I noted above that the contrast between universal biological law and historical accident is an unfortunate one. We should rather conceive of these models as testing the specific conditions are under which particular developmental, genetic, ecological, or evolutionary phenomena would arise. Under what circumstances could a third sex evolve? Under what circumstances could variation be directed rather than random? Under what circumstances could evolution be Lamarckian, with changes in phenotype causing changes in genotype? *Well-calibrated models* which showed the evolution of exotic possibilities—phenomena not observed in the natural world—would be very suggestive indeed. Unfortunately, to the best of my knowledge, there is little attempt yet to use A-Life techniques to model exotic life.

Most of the interesting work in A-Life falls into a third category. Much of the A-Life literature itself, and even more of the philosophical reflection on it, has focused on the issue of 'emergence' and 'self-organization'. This work links an empirical idea to a conceptual one. The central *empirical idea* here is that A-Life simulations show that surprisingly complex system-level

[2] Netspinner, a model of web evolution, is discussed extensively and approvingly in Dawkins [1996] in the course of a fairly sceptical discussion of evolutionary simulation. He argues that only in rather exceptional circumstances will simulations be robust.

behaviour arises out of locally interacting simple units. Complexly behaving systems require neither complex parts nor central direction. The elements in these models are quite simple units whose interactions are all governed by local rules; indeed, relatively simple local rules. But the behaviour of the system as a whole is often adaptively complex. Some social insect colonies provide natural examples of the phenomena in question: simply interacting simple creatures none the less produce complex, adaptive, and patterned behaviour at the macrolevel. So a good many of the most striking examples of A-Life models can be seen as undercutting the idea that fancy systems must be built of fancy components.

The *conceptual idea* is methodological. Since the interactions of the components determine system level behaviour we do not get much of a handle on what the system will be like by studying the components in isolation. Context is all. Emergence as an empirical phenomenon needs for its understanding new models of scientific explanation (see Burian and Richardson, Clark, and Hendriks-Jansen, all in Boden). It is worth noting that there is a serious tension between the pretensions of A-Life to be the science of universal biology, and its great emphasis on emergence. For the behaviour of emergent systems is radically unpredictable by any means other than simulation. In particular, analytic decomposition fails, for system level behaviour is generated by interactions between the elements. It is hard to reconcile the insistence on the unpredictability of the macrolevel and its dependence on fine-grained local interactions with the idea that A-Life is developing a universal biology. Emergence undercuts the prospects of universal biology. For system-level behaviour will typically be unpredictable. Moreover the macro/micro distinction is not absolute but domain relative. An organism's phenotype is a macrolevel phenomenon in developmental biology, but a microlevel one in ecology. So 'emergence-driven' unpredictability will trickle down.

A few examples might make these abstract points clearer. One good example is Reynolds' model of flocking behaviour (see Langton in Boden). He calls his simulated-creatures 'boids' and the rules they follow are very simple. Each acts

> to maintain a minimum distance from other objects in the environment, including other boids
>
> to match velocities with boids in its neighbourhood and
>
> to move towards the perceived centre of mass of the boids in its neighbourhood (Boden, p. 66).

Despite the simplicity of these units, boids simulate flocking rather well, flowing naturally around obstacles, and showing the illusion of co-ordination we see in schools of fish and flocks of birds. So this example shows how entities

following very simple, locally cued behavioural rules can form flocks whose global behaviour appears co-ordinated. Langton argues that the simulation of genotype/phenotype relations in A-Life models makes the same point. For 'genes' in these models are simple units with simple behavioural profiles responding to their local environment. Yet, jointly, in simulation as in life, these locally controlled interactions result in the development of complex and integrated phenotypes.

Some of these simulations are very suggestive. But what, exactly, do they show? What is their evidential status? This question is particularly important in thinking about Kauffman's work (Kauffman [1993]), discussed insightfully by Burian and Richardson (in Boden). For many see him as developing a picture of life that underplays the role of natural selection. Kauffman is often presented as *showing* both restrictions on the power of natural selection, and as *showing* that we do not need to invoke selection to explain 'order'. Order arises 'naturally'.

Kauffman's work exemplifies the theme we have just been discussing. Complex macroscopic organisation can derive from the interactions of simple systems under local control. However, the models are very abstract, and the key parameters very general. For Kauffman develops evolutionary models in which just two elements vary; N—the size of the population whose units vary in fitness (in these models often thought of as a population of genes)—and K, the 'connectedness' between members of N. The more other units each individual interacts with, the greater is K. So K measures the extent to which the fate of each unit is determined locally. As K goes up, local control goes down.

Kauffman derives some striking and lifelike general results from these models. These include the expectations that evolutionary radiations have a 'Cambrian-like' pattern. A lot happens fast, then not much happens at all. They include as well the idea that early ontogeny is more fixed than it is later because of the entrenchment of early mechanisms; that cell types increase as square root of gene number; and that cell types can switch at most only to a few other kinds.

But the most general and important result of the models that Kauffman explores is that connectedness damps down the effect of selection. Evolution under natural selection is possible only in a rather abstractly defined class of environments in which the number and linkage of the components is not too tight, and in which the fitness landscape is not too rugged:

> Kauffman says this reveals a 'fundamental restraint facing adaptive evolution'.... Everything depends on N and K. If K is much smaller than N, then there will be high optima, but any genotype will only be marginally better than its neighbours in the space of genotypes. If N is approximately the same order as K, then there will be a host of local optima, but any one

> genotype will only be marginally better than the average. In both cases, sub-optimal forms will be common (Boden, p. 155).

What are we to make of these results? They are clearly suggestive, but caution is called for. The very abstractness of the models makes their connection with the real biological phenomena difficult to evaluate. A mean-spirited approach would be to argue that these models are like simulations of host/parasite interactions that fail to show a co-evolutionary arms race. Consider, for example, Burian and Richardson's discussion (pp. 157–8) of Kauffman's idea that genotypes are self-organized, since given their degree of interconnection it's implausible to suggest that selection could prevent mutation and other disruptions 'spreading genotypes more evenly over the fitness landscape'. Yet genotypes are not just ordered and complex: they are very considerably differentiated from one another, and this in many ways must be the result of selection. The differences between primate genotypes may be partly drift, but surely many are the result of selection. So we *already know* selection can change genotypes, despite their apparently high connectedness. That cuts across the 'result' that as the 'connectedness' of a system goes up, and the number of elements in that system go up, then selection becomes increasingly ineffective. Kauffman's investigations into universal biology have discovered a 'constraint on selection' that shows that most actual biology is impossible, and much actual evolution has not happened.

I think there is a more generous way of thinking about these models. They lead us to ask how evolution under natural selection dodges the apparent constraints that would make it impossible. Is the number of effective units (the size of N) smaller than it would seem? Is effective connectivity less than it seems? It is often thought that the phenotypes of organisms must be 'modular', with some traits able to vary independently of one another, if natural selection is to be effective. So perhaps we should see these models as offering a hint that genotypes, too, are more modular than they seem.

More generally, we should treat A-Life models as 'how possibly' explanations. The foes of adaptationism have often made the point that we should not conflate how-possibly explanations with how-actually explanations, and that point is well taken.[3] Even so, how-possibly explanations are important. First, even in those areas where we think we have approximately the right how-actually story, an expansion of the space of possible explanations is often useful. For competing explanations indicate critical tests. How-possibly explanations are of course still more important when we deal with phenomena which are puzzling. A how-possibly explanation of, say, the evolution of human language would be useful because we have no good grip on what intermediate

[3] Though it is somewhat ironic than a good number of those on the anti-adaptationist bandwagon cite Kauffman and similar work in the same genre as offering reasons for scepticism about the power of natural selection, thus themselves conflating the possible with the actual.

forms of language would be like and why they were adaptive. Language poses a 'trajectory problem'.

There is a fairly unified body of A-Life that does contribute to our understanding in this way, because it does suggest possible solutions to a prima facie problem: the explanation of intelligent action in a complex and changing environment. Explaining intelligent behaviour has been seen as an intractable problem in AI, cognitive psychology and philosophy of mind because of the 'frame problem'. Intelligent behaviour seems to require the construction and use of an internal representation or model of the world, but updating and using such a model seems to require unreasonably great computational power (Fodor [1983]; Dennett [1984]; Pylyshyn [1987]). This problem has driven the development of a number of alternatives to 'classical AI' and A-Life has been the home of one of these. This issue is discussed by Kirsh, Clark, McFarland and Hendriks-Jansen (in Boden). In the Langton collection, the same issue is the focus of papers by Steels, Dyer, and Maes. Somewhat surprisingly, Brooks [1991] is not reprinted, though Kirsh's paper is a reply to him. Interestingly, these models involve a move away from merely informational existence: the aim is the construction of robots, not just programs; robots that interact successfully with their environments.

4 Situated agents

A-Life has seen the development of a number of anti-representational theories of action. On these views, intelligence is the result of the interaction between organism and environment. Developmental systems theorists deny that the gene code is preformed information that guides development. In a similar vein, these models deny that intelligent behaviour is (necessarily) guided by pre-existent representation and goal structures within the organism. Sometimes this line of thought suggests that the information to guide behaviour is constructed in behavioural interaction with the world, by organisms actively searching for specific, relevant cues. Sometimes the idea is that information is stored in the environment, not just the brain of the organism; ant pheromone trails obviously fit this picture. Agents that act intelligently in the world without benefit of a prior representation are 'situated agents'.

Kirsh takes the 'situated agent' program to consist of the following central ideas:

1. Behaviour can be partitioned into task-oriented skills. These are behavioural modules, each with distinct, hence parallel sensing/control requirements.
2. The behavioural repertoire of even complex creatures can be built from these modules, adding increments to a base of simple skills. Thus complex skills are combinations/sequences of these simpler ones.

3. 'Classical AI' has underestimated the available information in the environment that suffices to control these basic skills. These skills can be appropriately triggered and guided by local cues. 'World models' are almost always unnecessary. Intelligent behaviour does not need to be guided by information pre-stored in the organism.
4. Organisms co-ordinate their behaviour through a built-in preference structure using information that the environment provides.

> if all works well, the net effect is that as the world changes, either because the robot itself is moving through it, or because of external events, the robot will behave as of it is choosing between many goals. Sometimes it runs, sometimes it wanders, sometimes it feeds (Boden, p. 244).

The ambitious program is to suggest that models of 'situated agency' show that the 'frame problem' is a pseudo-problem. It depends on mistaken presuppositions about intelligent action. However, for the idea to be generally applicable, an organism must typically find itself in an environment that:

(1) provides the organism with specific and reliable cues. Each behavioural module must be tuned to a precise triggered stimulus, one highly correlated with the relevant distal feature of the environment, the 'affordance' relevant to the organism.
(2) provides the organism with local cues: 'only a fraction of the visible world must be canvassed to determine' the right action (Boden, p. 246). Situated agents escape the frame problem by avoiding having any *overall model* of their world, even an overall representation of their immediate environment. For that is how they escape the problem of update.

The Brooks bet is, then, that 'there is a reliable correlation between egocentrically noticeable properties of the environment and actions that are effective' (p. 249). So here once more we have the classic A-Life theme of emergence. Complex system-level behaviour (the intelligent, adaptive behaviour of organisms in their environment) emerges from locally governed interactions between relatively simple components. A similar methodological moral is drawn as well. There has always been a tradition within ethology of scepticism about laboratory-based, experimental studies of behaviour. If organisms are situated agents, storing some of their operational information in the environment, or developing it in interaction with very specific cues in their natural environment, caution about the experimental, manipulative approach makes sense.[4] We would need to be very cautious in extrapolating from laboratory

[4] As would scepticism about comparative psychology based on a few model organisms. The basic neural equipment—the basic internal mechanisms of control and learning—may well be fairly similar from organism to organism. But if the explanation of behaviour depends critically on the interaction between these intrinsic mechanisms and their environments, then neural similarity is no reason to expect overall similarity.

behaviour to natural behaviour and back again. To put it luridly, in shifting an organism from its natural environment we may be extracting part of the cognitive system, not the whole cognitive system. Furthermore, extracting a component is likely to be useful only if the behaviour of the system as a whole is not 'emergent'. A-Life theorists expect the interactions between the components, rather than the intrinsic features of the components themselves, to be most important for generating system-level behaviour.

This picture suggests the following research agenda, explored in various ways, and differing degrees of scepticism, by Clark, Kirsh, and Hendriks-Jansen. (i) How much adaptive behaviour is situated behaviour? (ii) How can the ideas behind situated agency be developed to explain more complex forms of intelligent behaviour? Kirsh is perhaps the most sceptical of this idea as a general theory of intelligent agency. Amongst his reservations is the problem of error. He points out that correcting errors and, especially, avoiding repeating them will often require some understanding of why action misfired. This is a special case, I suspect, of a more general phenomenon: feedback. An organism that can track an affordance only through a single, specific cue is very limited in its ability to use feedback to control and modulate its behaviour, for it is restricted to reliance on variation over time in that single cue. So if the cue can be misleading, or if there are time lags in response as the cue changes with the affordance (as there will be, for example, with chemical cues and other physical signals whose transmission speed is low relative to the action-time of the organism) feedback will be at best crude.

My own guess is that situated agency is a plausible picture of adaptive action only in benign, or at least not hostile, environments. I think it is no accident that in this program, A-Life models are of robots interacting with their physical surrounds—succeeding, for instance, in navigating their way around the walls of a room without having any overall representation of the room and its layout. Cue-driven behaviour offers a plausible picture of interaction with an indifferent physical environment. Thus we can conceive of, say, beaver dam repair as a sequence of skills driven by local cues. The sound of running water could initiate a random search along the inside of the dam walls; the feel of the current near a break could then induce a local search, then simple behavioural rules could guide action in dam repair. The beaver need not, for this task, have any overall model of the dam and its state. Equally, cues can be sufficient to drive behaviour in co-operative interaction, where each organism is trying to make its intentions as explicit as possible. Ants communicate with one another by producing local cues. But much animal behaviour takes place in a hostile world of predation and competition. Predation is not just a danger to life and limb; predation results in epistemic pollution. (Prey, too, pollute the epistemic environment of their predators.) Hiding, camouflage, and mimicry all complicate animals' decision problems.

For cue-driven behaviour to be adaptive, the cue itself must be detectable and discriminable, though perhaps it requires the organism to probe its environment. Further, there must be a stable cue-affordance relationship. What the cue tells the organism about its world needs to be (fairly) statistically independent of what else is in the local scene. That is why the organism does not need to represent that local scene as a whole. To put it another way, the organism's local environments must be homogenous with respect to the cue-affordance relationship. Hostility imposes a cost on probing. It imposes a cost on animal action taken to disambiguate a cue, or to locate one ('If called by a panther, don't anther'). Second, it makes local environments heterogeneous with respect to easily discriminated cues and the affordances they signal. Deceptive fireflies mimicking the female signal to the male decrease the overall reliability of the signal-mate relationship, as the firefly environment becomes heterogeneous with respect to the 'species-specific' signal.

5 Conclusion

I hope this notice gives a feel for the variety and heterogeneity of A-Life. There is lots not to like in this field: hype and exaggeration all too reminiscent of the early days of AI; and, more harmlessly, gee-whiz technophilia as well. But there are interesting and provoking ideas as well. The collections themselves reflect this diversity. For philosophers interested in the philosophical and theoretical upshot, rather than the models themselves, the Boden collection is clearly the best. Those with more technical interests will prefer the Langton collection though I suspect (though I am no expert) that real insiders will find some of the papers too skimpy on detail and insufficiently new. The Langton collection may be most useful to technically minded outsiders wanting a briefing on what has been going on.

Department of Philosophy
Victoria University of Wellington

References

Brooks, R. A. [1991]: 'Intelligence Without Representation', *Artificial Intelligence*, **47**, pp. 139–59.

Dawkins, R. [1996]: *Climbing Mount Improbable*, New York, W. W. Norton.

Dennett, D. C. [1984]: 'Cognitive Wheels: The Frame Problem of AI', in C. Hookway (*ed.*), *Minds, Machines, and Evolution*, Cambridge, Cambridge University Press, pp. 129–52.

Dennett, D. C. [1995]: *Darwin's Dangerous Idea*, New York, Simon & Shuster.

Fodor, J. A. [1983]: *The Modularity of Mind*, Cambridge, MA, MIT Press.

Griffiths, P. and Gray, R. [1994]: 'Developmental Systems and Evolutionary Explanation', *Journal of Philosophy*, **XCI**, 6, pp. 277–304.

Kauffman, S. A. [1993]: *The Origins of Order: Self-Organisation and Selection in Evolution*, New York, Oxford University Press.

Lycan, W. G. [1990]: 'The Continuity of Levels of Nature', in W. G. Lycan (*ed.*), *Mind and Cognition*, Oxford, Blackwell, pp. 77–96.

Nilsson, D.-E. and Pelger, S. [1994]: 'A Pessimistic Estimate of the Time Required for an Eye to Evolve', *Proceedings of the Royal Society of London,* **B 256**, pp. 53–58.

Oyama, S. [1985]: *The Ontogeny of Information*, Cambridge, Cambridge University Press.

Pylyshyn, Z. [1984]: *Computation and Cognition*, Cambridge, MA, MIT Press.

Pylyshyn, Z. (*ed.*) [1987]: *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Norwood, NJ, Ablex.