OXFORD READINGS IN PHILOSOPHY

The Philosophy of Artificial Life

edited by
Margaret A. Boden

INTRODUCTION

THE INTELLECTUAL CONTEXT OF ARTIFICIAL LIFE

Artificial Life (A-Life) uses informational concepts and computer modelling to study life in general, and terrestrial life in particular. It raises many philosophical problems, including the nature of life itself.

There is no universally agreed definition of life. The concept covers a cluster of properties, most of which are themselves philosophically problematic: self-organization, emergence, autonomy, growth, development, reproduction, evolution, adaptation, responsiveness, and metabolism.

Theorists differ about the relative importance of these properties, although it is generally agreed that the possession of most (not necessarily all) of them suffices for something to be regarded as alive. It is not even obvious that, as A-Life scientists assume, life is a natural kind. In other words, 'life' may not be a scientifically grounded category (such as water, or tiger), whose real properties unify and underlie the similarities observed in all those things we call 'alive'. Instead, it may pick out a rag-bag of items with no fundamental unity. Advances in A-Life should help to resolve this issue, and to identify the underlying unities (if any).

The field raises fundamental problems also about the nature of explanation in various areas of biology, and in cognitive science. One of these is whether (and if so, how) psychological explanations should be constrained by biological facts. Another concerns the choice between computational and dynamical explanations. Some A-Life researchers argue that dynamical systems theory is more promising than computational theories in investigating life and mind. A dynamical system is one whose states change over time according to some rule. Computational systems are a special case, whose states change stepwise according to the instructions in the program. Many dynamical systems—the weather, for example—change continuously, and are described by differential equations such as those used in physics. Which type of change is characteristic of cognitive systems is a controversial question. Such dynamical systems may be closely coupled, changes in one affecting general parameters (not just individual states) in the other, so that it is difficult to make a principled distinction between the 'two' systems. So, for instance, some A-Life workers regard the distinction between organism and environment as deeply problematic.

Moreover, since most (though not all) A-Lifers define life in informational terms, A-Life places functionalism in a new scientific context. Functionalism is the philosophy of mind which defines mental states in terms of their causal relations with other mental (and environmental) states, and which assumes that these causal relations are, in principle, expressible in computational terms (Putnam 1960, 1967). In short, pain, fear, or jealousy are to be defined not as phenomenal experiences, nor as physical events in the brain, but as abstract functional (causal) roles. Functionalism (with respect to mind) is widely accepted within AI, most of whose practitioners define mental phenomena in informational terms. Many A-Life workers take a similar view of life itself.

A-Life is highly interdisciplinary (Langton 1995). It has profuse intellectual connections with theoretical biology, including morphology, embryology, ethology, evolutionary theory, and ecology. It has links with psychological sciences such as developmental and cognitive psychology, computational neuroscience, neuromorphic engineering (wherein brainlike systems are built using real neurones), and computational psychology. And it is related to artificial intelligence (AI) and cognitive science, which try to model psychology much as work in A-Life attempts to illuminate biology.

It is related also to biochemistry, economics, and the mathematics of complexity. However, the overall emphasis of this collection concerns A-Life's relevance to theoretical biology and cognitive science.

Commercial applications of A-Life range widely: from pharmaceutical research, through financial services and telecommunications, to the visual arts. The bats in the film *Batman Returns* were controlled by simple A-Life flocking algorithms (Reynolds 1987), and sprouting vegetation and 'natural' surface-patterns can be automatically generated by algorithms defining branching and pigmentation (Prusinkiewicz 1994). Evolutionary programming is being used in computer art (Todd and Latham 1992; Sims 1991), behavioural animation (Reynolds 1994*a*, 1994*b*; Sims 1994), and architectural design (Frazer 1995).

Clearly, then, A-Life is relevant to diverse philosophical interests. So far as possible, the papers in this volume are grouped around specific themes. Chapter 1 provides a general overview of the field. Chapter 2 compares A-Life with AI, focusing on the concept of autonomy. Chapters 3 to 6, and part of Chapter 7, describe examples of A-Life research, emphasizing its relation to biology and cognitive science. Chapters 7 to 10 discuss various explanatory strategies in A-Life, and relate them to approaches in AI and cognitive science. Chapters 11 to 13 focus on the concept of life in general. A-Life's relation to functionalism, and the feasibility of 'strong' A-Life, are explored in Chapters 14 and 15.

A-LIFE AND AI

The central concept of A-Life, excepting *life* itself, is *self-organization*. Self-organization involves the emergence (and maintenance) of order, or complexity, out of an origin that is ordered to a lesser degree. That is, it concerns not mere superficial change, but fundamental structural development. This development is 'spontaneous', or 'autonomous', following from the intrinsic character of the system itself (often, in interaction with the environment) instead of being imposed on the system by some external designer. In that sense, A-Life is opposed to classical AI, in which programmers impose order on general-purpose machines (Boden 1987).

AI (the subject of another volume in this series: Boden 1990) is the attempt to build and/or program computers to do the sorts of things that minds can do—sometimes, though not necessarily, in the same sort of way. Some AI researchers approach their work as engineers: their goal is to achieve some technological task, irrespective of how living creatures achieve it. But others aim to throw light on the principles of intelligence in general, or of human psychology in particular. They accordingly try to make their models work in ways analogous to biologically based minds.

Some of the things that minds can do are normally regarded as involving intelligence: for example, playing chess, or making a medical diagnosis. Others are not: seeing, visuomotor coordination, perhaps even speaking one's native language. Early AI focused on the more 'intelligent' tasks, recognizing only later that apparently simple abilities, many of which we share with some animals (and some of which are modelled also in A-Life), are even more difficult to model than is highly educated human expertise.

Some such abilities seem to require a form of computer modelling—'connectionism'—very different from step-by-step 'classical' AI-programming. Connectionist systems consist of networks of simple interconnected units, wherein concepts may be represented as an overall pattern of excitation distributed across an entire network. These networks are parallel-processing systems, in the sense that all the units function (exciting or inhibiting their immediate neighbours) simultaneously. Because they are broadly inspired by the neurones in the brain, connectionist models are sometimes called neural networks. (Philosophical problems associated with classical and connectionist AI are explored in another volume in this series (Boden 1990).)

A-Life and AI differ significantly, despite having strong historical and methodological links. In classical AI, one starts with a general-purpose (task-neutral) machine, and attempts to write a program to make the machine perform the task required. In A-Life, by contrast, the aim is to define simple reflex-like rules from which the more complex target-behaviour will emerge (Braitenberg 1984). Moreover, A-Life avoids the explanatory emphasis on 'representation' found in most AI, connectionism included (Chs. 7–10).

Another difference is that A-Life eschews classical AI's emphasis on top-down processing, focusing instead on simple processes that work bottom-up to generate order at a higher level. In top-down processing, a high-level representation of the task or sub-task (a goal, grammar, or expectation) is used to initiate, monitor, and/or guide detailed actions. In bottom-up processing, it is the detailed input of the system which determines what will happen next. Classical AI uses both approaches, but top-down processing is the more common.

Connectionist AI works bottom-up in that the behaviour of a connectionist model depends on the local interactions of the individual units, none of which has an overall view of the task. But even bottom-up connectionism differs from A-Life, for connectionist systems cannot (yet) develop their own principles of organization. They can learn to settle into particular equilibrium-states, by changing myriad connection-weights. Their fundamental organization, however, remains unchanged. Nor can connectionist AI model multi-level order, a prime interest of A-Life. 'Recurrent' networks, in which the output of one layer of units can be fed back as input to a preceding layer, can approximate hierarchy to some degree. But hierarchy is better modelled by classical AI.

Not surprisingly, then, some (though not all) A-Lifers explicitly distance themselves from AI. A participant in the first A-Life conference acknowledged that 'artificial life studies have closer roots in artificial intelligence and computational modelling than in biology itself', but went on to attack overly computational approaches (Ch. 15). And many who work with embodied creatures (natural or artificial), as opposed to the disembodied systems typical of AI, stress the differences between their approach and AI (Chs. 7 and 10).

However, the differences between A-Life and AI should not be exaggerated. There is significant conceptual overlap between the two fields (some types of research are published in both A-Life and AI journals). For example, certain guiding ideas of A-Life also inform AI work in situated robotics and animate vision (jointly named *nouvelle* AI), and in evolutionary robotics.

Situated robots react directly (bottom-up) to environmental cues rather than (top-down) to internal world-models or representations (Chs. 7 and 8). They are typically described as autonomous, another concept central to A-Life. This usage is not entirely felicitous: situated robots do satisfy some criteria of autonomy, but the reflexive self-direction of much human action is better modelled by classical AI (Ch. 2). However, the emphasis on self-directed control rather than outside intervention is shared with A-Life.

AI research on animate vision is related to A-Life also, for it stresses the role of the creature's own bodily movements in affecting the perceptual feedback used, in turn, to direct future action (Ballard 1991).

Evolutionary robotics is even closer to A-Life. Indeed, it falls equally into A-Life and AI, even though leading practitioners favour dynamical systems theory rather than symbolic computation in describing their work (Ch. 7). Here, the robot's 'brain' and/or sensory-motor morphology are not specifically designed by the roboticist, but are automatically evolved within its task-environment (Cliff, Harvey, and Husbands 1993; Husbands, Harvey, and Cliff 1995; Thompson 1995).

The philosophical similarity of the two fields is disputed. Some A-Life sympathizers refuse to apply ideas from the philosophy of AI to A-Life (Keeley 1994). Others favour philosophical assumptions fundamentally opposed to those of AI (Ch. 7). However, there are grounds for assimilation, too. If one assumes that intelligence is grounded in life, then AI is somehow continuous with A-Life (Ch. 12). Largely parallel arguments arise about the adequacy of a purely abstract functionalism to describe life and mind (Chs. 14 and 15). To understand human minds, we need explanations drawn from both (and from biology and neuroscience), with an appreciation of how they fit together (Ch. 9).

ORIGINS OF A-LIFE

Historically, it is indisputable that A-Life and AI are close cousins. Although AI flourished before A-Life, their intellectual seeds were sown at much the same time, by some of the same people. Their common intellectual ancestors—apart from the very early machine modellers (see Ch. 1)—were Alan Turing and John von Neumann.

Turing initiated computer science in the 1930s, outlined the agenda of AI in the 1940s, and pioneered the design of working computers (for code-cracking) in World War II. As if all that were not enough, in the early 1950s he published a mathematical paper on morphogenesis (the development of biological form), whose implications are still being explored (Turing 1952).

He proved that relatively simple chemical processes (described in abstract mathematical terms) could generate new order from homogeneous tissue. Two or more chemicals diffusing at different rates could produce 'waves' of differential concentrations which, in an embryo or growing organism, might later prompt the repetition of structures such as tentacles, leaf-buds, or segments. Diffusion waves could engender ordered cell-differentiation in one, two, or three dimensions. In 3D, they could, for instance, cause embryonic gastrulation, in which a sphere of homogeneous cells develops a hollow (which eventually becomes a tube).

Modern embryology and morphology owe much to Turing's inspiration. Recently, his own equations have been used to generate spot-patterns and reticulations like those of dalmatians, cheetahs, leopards, and giraffes (Turk 1991). This work requires powerful computers, to vary the numerical parameters in Turing's differential equations, calculate the results, (often) apply two or more equations successively, and express all these numerical results in graphical form. Turing himself had only primitive computers to help him. Because (as he remarked) much better machines were needed to follow up his ideas, his paper—though recognized as an important contribution to analytic biology—did not immediately spawn A-Life as a computational discipline.

Von Neumann, too, pioneered the artificial sciences. His design of the digital computer in the late 1940s was partly inspired by early ideas in computational neuroscience (McCulloch and Pitts 1943), and in turn enabled AI experiments to start in the 1950s (Feigenbaum and Feldman 1963). At much the same time, he was doing pioneering theoretical research on cellular automata, including self-reproducing systems (Burks 1970). This was A-Life in embryo.

A cellular automaton is a computational 'space' made up of many cells. Each cell changes according to the same set of rules, taking into account the states of neighbouring cells. The system moves in time-steps, all the cells normally changing—or not—together. (Asynchronous systems are also possible: identical rule-sets may give very different results if the cells have independent 'clocks' (Ingerson and Buvel 1984; Bersini and Detours 1994).) After each global change, the rules are applied again.

Von Neumann was interested in the order spontaneously generated in cellular automata following simple rules. (Turing's paper on diffusion gradients used similar ideas, for instance describing the emergence of differential concentrations within a twenty-cell ring.)

As part of his work on cellular automata, von Neumann studied the 'logic' of reproduction. Having already shown that a self-replicating *physical* automaton is in principle possible, he defined the *functional* (computational) features of reproduction. Several years before the discovery of DNA and the genetic code, he realized that part of the self-replicating system must function both as instructions and as data (a self-description). He even defined a universal replicator, a computational system capable of reproducing any system. (Biological species are not universal replicators: cats give birth only to kittens.) He also remarked that errors in copying the self-description could lead to evolution, which might thus be studied computationally.

It is now known that cellular automata are in principle equivalent to Turing machines (Langton 1986). (Turing machines are theoretically defined computing systems with an infinite tape, capable in principle of performing any possible computation; actual computers are approximations to Turing machines (Turing 1936).) However, von Neumann's A-Life ideas, like Turing's work on morphogenesis, were neglected for many years by computer modellers. Their attention focused instead on AI and cybernetics, which could advance with the help of early computer technology. Von Neumann's A-Life research (like Turing's) was purely theoretical, because to explore its implications required considerable computer power.

After von Neumann's death in 1956, mathematical work on cellular automata continued, and even hit the popular scientific press (Gardner 1970). But it was unavoidably limited until recently because of the lack of computational power.

As for computational evolution (which von Neumann foresaw), this took some time to develop. It had to await John Holland's formal definition of genetic algorithms (GAs) in the 1960s (Holland 1975); the actual implementation of GAs (for modelling optimization) in the 1980s (Holland *et al.* 1986); and their extension from fixed- to variable-length descriptions (for modelling species-evolution) in the early 1990s (Harvey 1992). Genetic algorithms are inspired by the genetic mutations and (especially) cross-over found in biology, and are widely used in A-Life and AI. They have been applied in work on induction and other aspects of the philosophy of science (Holland *et al.* 1986), but here they are of interest primarily for the light they can shed on the general principles underlying evolution. (Genetic algorithms are briefly described in Chs. 1 and 2, and discussed in more detail in Chs. 3 and 5).

In sum, the central ideas of A-Life originated in mid-century. But, largely because of its need for high-performance computing, the field achieved visibility only in the 1990s. For four decades, the scattered examples of A-Life research being done were not even recognized as a unitary endeavour. The unification, in so far as a field so diverse can be so described, occurred in the late 1980s, when the term 'A-Life' was coined.

A-LIFE'S AGENDA

A-Life was christened at a conference (in 1987) to which researchers in many different disciplines were invited. The paper that named the field, unifying diverse research under the one label, is reprinted here as Chapter 1. Christopher Langton's 'Artificial Life' is widely regarded as the manifesto defining A-Life's agenda. It is partly historical, partly an overview of research, and partly a statement of the aims and presuppositions of the field.

Anyone who expects A-Life's manifesto to begin with a definition of life will be disappointed. Langton argues that one cannot give necessary and sufficient conditions for anything on the basis of only one example. All the living things we know of share the same basic biochemistry, and a common genetic descent (the DNA code is biologically universal). Moreover, they have all evolved in response to local historical accidents, so that general principles are difficult to identify. Even if we could do this for terrestrial life (as A-Life may help us to do), we could not be sure that all living things must have the same properties. But A-Life might enable us to synthesize new systems recognizable as alternate life-forms. We could then distinguish 'life-as-it-is' from 'life-as-it-could-be'. In short, Langton sees A-Life as a way of discovering *what life is*, as well as *how life is possible*.

Langton sees life as an abstract phenomenon, a set of vital functions implementable in various material bases. Life consists of dynamic processes, organized in certain ways. Among the general principles he outlines in Chapter 1 are self-organization, self-replication, emergence, evolution, and the (epigenetic) unpredictability-gap between genotype and phenotype. (The genotype is the set of genes, or hereditary material, from which the living system develops; the phenotype is the actual form of the mature individual.)

In addition, Langton discusses many aspects of information processing, which he—like von Neumann—sees as fundamental to life. This is no mere flight of hermeneutic fancy. For Langton, *information*, *communication*, and *interpretation* are real computational properties of certain formally describable systems. These assumptions are widespread in the field. One philosopher of A-Life has even described life as 'matter with meaning' (Pattee 1995).

One does not have to be a devotee of Langton, or even of A-Life, to share this view. Informational (intentional) concepts were widely used by theoretical biologists long before Langton gave A-Life its name (Boden 1980). Familiar examples include the genetic 'code', with its attendant 'reading', 'interpreting', and 'transcription'; 'programs' of 'instructions' conveyed by specific cells or chemicals; 'messages' passed within and between cells; and cybernetic notions of 'informational feedback' and oscillatory 'control'.

One reason for the use of informational concepts in biology is that many biological processes are arbitrary with respect to biochemistry. The same process (defined at the biological level) may be instantiated by different mechanisms in different species: like mental processes as defined by functionalism, the process shows multiple realizability. It is therefore useful to define biological processes at an informational level, independent of the underlying biochemistry.

Moreover, biochemistry does not map on to biological function in any simple fashion, since the 'meaning' of a given metabolite (some molecule produced by biochemical processes within living cells) can vary. Indeed, a metabolite may have a different significance not only across species, but also at distinct phases of the life cycle of a single species. In slime moulds, for instance, one and the same chemical substance (cAMP) has different biological functions, or meanings, in the three phases of the organism's life-history. In the first phase, cAMP

causes separated amoebic cells to aggregate into a mass, or slug. In the next, it acts as the pacemaker for the contractions by which the slug moves along the ground. Finally, it causes the homogeneous cells of the slug to differentiate into three distinct types, forming a base, a stalk, and a spore-containing head. The three developmental contexts thus result in differing 'interpretations' of the cAMP 'messages' (Boden 1980).

In using informational language to describe biological processes, one has to have some way of identifying what the relevant information is. The meanings of biological 'codes' or 'instructions' are determined by their causal powers and evolutionary history. Likewise, a feature within an evolving A-Life system draws its significance from its contribution to fitness over many generations. Philosophies that appeal to evolutionary history in explaining intentionality (e.g. Millikan 1984) are broadly consonant with A-Life.

Just which informational concepts are likely to be of most use to biologists is controversial. Linguistic and programming metaphors are still widespread in theoretical biology. But some A-Life researchers reject them in favour of cybernetic and informational concepts underpinned by dynamical systems theory. Which cognitive-informational concepts are most useful for biology and A-Life is an empirical question, although it may be linked to wider philosophical claims (see Ch. 7).

Langton describes A-Life as 'unabashedly' reductionist, because it holds that high-level phenomena depend on simple interactions between lower-level processes. But (like psychological functionalism) it is also anti-reductionist, in that it treats those phenomena as real properties, described by ineliminable theoretical concepts. The behaviour of flocks of birds, for instance, must be described on its own level, although it results from the behaviour of individual birds. Similarly, the information-bearing functions of DNA base-triplets are real properties of the genetic system, not expressible in terms of—though explicable by means of—biochemistry. Such higher-level properties are often described as emergent (although the strong definition of emergence in Chapter 9 excludes some cases discussed by Langton).

An example of what might be involved in 'discovering *what life is*' is given by Langton in a later paper (Langton 1991), in which he defines a measurable property that he claims is a necessary condition for life, wheresoever in the universe it may occur. Information-processing, and life, require a certain type of complexity. The system must be dynamic, yet novel patterns must be reliably related to their predecessors. A qualitative distinction between four types of complexity, only one of which can support information-processing, already existed (Wolfram 1984). But Langton's was the first quantitative definition of order/disorder. In extensive experiments with computer models of cellular automata, he matched the relevant type of complexity with the numerical values of his statistical measure. He found that 'interesting' complexity arises only within a very narrow range. Only then is life possible. Outside those values, the system's behaviour is boringly homogeneous, rigidly periodic, or uselessly chaotic.

Most descriptions of A-Life, Langton's included, emphasize the importance of autonomy in living systems. This concept is explored in my own paper on 'Autonomy and Artificiality' (Ch. 2). The sciences of the artificial support two opposing intuitions concerning autonomy. One, characteristic of classical AI, is that determination of behaviour by the external environment lessens an agent's autonomy. The other, characteristic of A-Life and situated robotics, is that to follow a preconceived internal plan is to be a mere puppet.

These intuitions can be reconciled, since autonomy is not an all-or-none property. Three dimensions of behavioural control are crucial. First is the extent to which response to the environment is direct (determined only by the present state in the external world) or indirect (mediated by internal mechanisms partly dependent on the creature's previous history). Second is the extent to which the controlling mechanisms were self-generated rather than externally imposed. And third is the extent to which internal directing mechanisms can be reflected upon, and/or selectively modified. Autonomy is the greater, the more behaviour is directed by self-generated (and idiosyncratic) internal mechanisms, nicely responsive to the specific problem-situation, yet reflexively modifiable by wider concerns.

Some senses of autonomy are indeed illuminated by work in A-Life. However, the strongest sense—human freedom—is more nearly approached by theories based on classical AI. Self-reflection, deliberation, and reasoned prioritizing are closer to classical AI architectures than to the direct environmental embedding typical of A-Life and *nouvelle* AI.

It does not follow that the artificial sciences must have a dehumanizing effect, undermining our confidence in our self-control. Far from denying the reality of what we call free choice, they help us to appreciate its complexities and to understand how it is possible.

BIOLOGICALLY RELEVANT A-LIFE

Chapter 3 is 'An Approach to the Synthesis of Life', by Thomas Ray. Ray is a tropical botanist and forest conservationist, whose A-Life work simulates evolution and co-evolution. Biological species evolve largely by adapting to the presence of other species. So do the 'species' in Ray's computer model (see also Hillis 1991). Simulated prey evolves more quickly in the presence of predators, and simulated hosts become infested with (and sometimes resistant to) parasites, which force the host to replicate the parasite rather than itself—and the parasites, in turn, are plagued by hyper-parasites.

Ray's 'digital organisms' are sequences of machine instructions, running on Tierra. (Tierra is a virtual computer simulated on a real computer, because Ray's creatures might otherwise pass from one real computer to others, infecting them all with proliferating code.) The original seed is a self-replicating program, many of whose descendants will be self-replicators also. The organisms compete for Tierra's CPU-time and memory-space, much as terrestrial organisms compete for energy and geographical space.

The instructions are sometimes executed imperfectly (an instruction to shift bits to the left may move them to the right instead); and the background operating system randomly flips bits during 'rest' and replication. This imperfect execution, maintenance, and copying is analogous to mutation. Tierra's memory is prevented from filling up by a computerized analogue of death, culling (usually) instruction-sequences that are old and/or have led to errors.

The single 'ancestor' is an eighty-instruction sequence, containing the code for self-replication. When self-replication works imperfectly, a daughter-sequence may contain no copy-procedure. Some such daughters are utterly sterile, but some may induce other instruction-sets to make grand-daughters in their own image. This parasitic behaviour is possible because, although no Tierran creature can change the code (genome) of another, it may read and even execute parts of another's code—for example, borrowing a neighbour's copying powers to produce copies of itself.

If the parasitic species (lacking the copy-procedure) is shorter than the host species, it occupies less memory-space, so tends to increase relative to the host. Eventually, if the parasite is to survive, an ecological balance must be found; for the parasite cannot reproduce without some host. Ray finds that after thousands of computer-generations there is usually a variety of species, of different sizes and longevity, and different ecological relations (independent, parasitic, symbiotic, and so on).

Evolution is grounded in complex probability distributions, which in Tierra can be precisely varied and the consequences compared—so providing an experimental medium for quantitative studies of evolution. Species can evolve more quickly in the presence of predators because the predator displaces the prey from sub-optimal 'solutions', or local maxima. (Compare a group of short-sighted mountaineers, standing on what they mistakenly believe is the summit: they could reach the true summit only if something were to remove them from the local peak, for then they might start climbing at the base of the summit.) But what difference does it make if the local maxima are many or few, grouped or scattered, steep or gently sloping? Such questions can be addressed computationally by comparing evolution within different fitness landscapes. Similar A-Life research has explored the effects of various mutation rates and population sizes, and of differing degrees of interaction between populations (Kauffman and Johnson 1991).

Also of interest is the occurrence of evolutionary 'leaps' in Tierra, wherein sudden change is observed after many generations of quiescence. This recalls the theoretical dispute between gradualism and saltationism (Eldredge and Gould 1972). Ray shows that the quiescence is merely apparent: outwardly invisible genetic changes gradually accumulate until reaching a 'critical mass', when they cooperate to cause phenotypic change.

Ray's ultimate aim, like Langton's, is the production of new life. Indeed, his preferred term for A-Life is 'Synthetic Biology'. He is not thinking only of the synthesis of new carbon-based organisms, or novel biochemistries (a prominent research theme in A-Life). For Ray, biochemical metabolism is inessential. He defines a living system as one which is 'self-replicating, and capable of open-ended evolution'. These criteria are met by the 'organisms' described in Chapter 3, and by those being seeded in the *Digital Reserve* spread across the Internet (Ray 1994). Accordingly, Ray claims these computer-species are literally alive. Even computer viruses, if they can evolve, are alive—or as alive as real viruses are (cf. Spafford 1991, 1994). (Because viruses depend on non-viral cells for reproduction, many biologists deny they are alive.)

Such claims are highly controversial. Many people would endorse Ray's work as a useful exercise in theoretical biology, without accepting his robust philosophical interpretation of it. Borrowing the terminology John Searle (1980) uses in discussing AI, they would say that even if weak A-Life is scientifically valuable, strong A-Life is impossible. As later chapters show, however, proving that Ray's concept of life is inadequate is not easy.

Scepticism about the synthesis of new life-forms may be based in the assumption that the origin of biological life, if not literally miraculous, was almost inconceivably improbable. Many precise details, it seems, had to be fulfilled for life to emerge. Since the odds against this are astronomical, life may even be a unique cosmic accident, occurring only on Earth. The A-Life research discussed by Richard Burian and Robert Richardson in 'Form and Order in Evolutionary Biology' (Ch. 4) suggests that this scepticism is misplaced. Far from being unique and unpredictable, the emergence of order, and even of self-reproduction, is virtually inevitable in systems of a certain degree of complexity.

To study the spontaneous origin of order, Stuart Kauffman (1993) considers information-processing systems made up of many interacting units. He interprets these abstractly defined systems in terms of vital phenomena on various levels: auto-catalytic molecules generating a connected metabolism, genes, simple neural networks, ant colonies, ecosystems, and market economies.

Like cellular automata, which they closely resemble, Kauffman's self-organizing networks work in global 'time-steps'. He asks what will happen if each unit responds in a given way to the influence of two or more others, chosen at random from those available. The units and interactions are very simple. A small set of logical functions (for instance, 'on' and 'off') may be randomly assigned to each one of a thousand units, with simple rules (also chosen at random) specifying what interactions will lead the unit to make which response. From such seemingly unpromising beginnings, Kauffman uses probability theory to show that dynamic order, of various generic types, is bound to arise eventually—where 'eventually' may be a surprisingly brief period.

To what extent these intriguing mathematical results apply to advanced biological systems is still unclear. For instance, Kauffman identifies systematic relationships between the number of units and the number of types of order that emerge. He suggests that this explains why creatures with larger genomes tend to have a greater variety of cell-types. However, some species have much more DNA than closely comparable species do; only if the genome is largely inert or 'junk' DNA does this fit Kauffman's picture. Again, he suggests that networks whose units connect up with more than five others will develop chaotic, not ordered, dynamics. Yet a human neurone may be interconnected with many thousands of others: how, then, can the brain be a superb information-processor?

If Kauffman's work does capture the essence of biological form, it has two important implications. First, the potential for generating life is an inherent property of matter. Kauffman's statistical arguments are drawn from physics. and he sees biological order as a natural consequence of basic physical laws. Second, genetics (molecular biology) and natural selection are not the only explanatory principles in biology, as neo-Darwinism assumes (Dawkins 1976, 1986). Granted, the genes specify the initial conditions (at various stages) of development, so crucially affect the resulting organism. And natural selection determines which of the slightly differing variants of developed forms will thrive. But, if Kauffman is right, these two factors alone cannot explain the form of living things. That an ordered organism develops at all, out of relatively homogeneous, unstructured beginnings, is due to the inherent self-organizing properties of complex systems (Depew and Weber 1994, chs. 15–16). In Kauffman's words, 'Evolution is not just "chance caught on the wing". It is not just a tinkering of the ad hoc, of bricolage, of contraption. It is emergent order honoured and honed by selection' (1993: 644).

A complete theoretical biology would thus require the integration of various kinds of explanation. Abstract analyses of generic form cannot tell us what the detailed mechanisms are which underlie actual metabolisms and morphologies. Those are questions for biochemistry, physiology, and genetics. Nor can they tell us what historical contingencies have shaped morphological change through natural selection. However, studies of the origin of order in the absence of natural selection may help us to understand, and even to measure, the mutual interactions of selection and self-organized form.

In 'Evolution—Natural and Artificial' (Ch. 5), the biologist John Maynard Smith asks whether A-Life might help us understand evolution. Less enthusiastic than Richard Dawkins, who says '[As] a dyed-in-the-wool, radical neo-Darwinist . . . I really have been led to think differently as a result of creating, and using, computer models of artificial life which, on the face of it, owe more to the imagination than to real biology' (Dawkins 1989: 201), Maynard Smith insists that many questions about the nature of evolution call for more biological data, not more theoretical models. Moreover, he sees computational models based on actual genetic systems as generally more useful to biologists than simulations of evolution in general.

For one class of questions, however, he grants that the latter approach may be helpful. Since biologists know of only one case of evolution (all terrestrial life appears to have the same origin), it is difficult to answer questions about evolution in general. For example: Has there been time for natural selection to produce complex living creatures? Is the adaptive landscape smooth, or rugged? And what features of terrestrial genetic systems are necessary, rather than historically contingent? The specific genetic code, for example, may not be a necessary feature of biochemical evolution. But Maynard Smith suggests that there must be some digital code involved, and that it cannot support (Lamarckian) inheritance of acquired characteristics. A-Life might confirm those suggestions, and might identify further necessary conditions of evolution as such.

Maynard Smith's insistence that A-Life models be firmly grounded in biological realities must be heeded, if A-Life is not to descend into ignorant dilettantism or mere playing around with computers. The sloppiness of much early AI reduced its interest for psychology and even for 'pure' technology (McDermott 1976), and some current A-Life has been similarly criticized (Miller 1995). Its biological potential is greatest in areas unamenable to the mathematical modelling and traditional computer simulation already widespread in biology. A-Life techniques (unlike these earlier methods) can model systems in which the component units are not simple, homogeneous, and predictable. Consequently, it offers a rich experimental and analytic context for investigating biological processes on several levels of complexity and over differing time-scales.

For example, work on the co-evolution of pursuit and evasion uses evolutionary simulation methods to track the emergence both of complex inter-agent behaviours and of the underlying sensory, neural, and motor systems. This work also enables one to compare the effects of varying the relevant factors in a systematic and measurable way (Cliff and Miller forthcoming; Miller and Cliff forthcoming).

Much A-Life work focuses on problems studied in ethology. This may involve the computer modelling of actual neurophysiological mechanisms (Cliff 1991; Webb 1994), or the construction of mobile or simulated artificial animals, or 'animats' (Mayer and Guillot 1991, 1994; Wilson 1991). The environments of animats are simpler than the real world, but typically involve a significant degree of unpredictability and 'threat'. This research may deal with single organisms, or with 'social' behaviour within groups of animals. Some A-Life projects consider specific behaviours of identified species, while others address cross-species behaviours such as flocking (Reynolds 1987), fighting (Sims 1994), pursuit/evasion (Cliff and Miller forthcoming; Miller and Cliff forthcoming), or communicative strategies (De Bourcier and Wheeler 1995).

The behaviour of animals is viewed on an even more general level in Chapter 6. In 'Animals as Cost-Based Robots', David McFarland explains all animals' activities in terms of cost-functions defined relative to their environments. This concept, he suggests, should be applied also to robots.

McFarland argues that animals do their 'decision-making' in a way fundamentally different from the internal world-modelling of traditional robotics. They thus avoid the frame problem: how to plan future action without considering a host of irrelevant details, and without being obstructed during execution by unanticipated consequences (Boden 1990, chs. 7–9). Classical AI depends on general-purpose computers, but there are no general-purpose animals: each species behaves appropriately only within its own ecological niche. Appropriate behaviour depends not on detailed planning, but on relatively direct reactions to environmental (and internal bodily) factors. The simplest animals are probably pure reflex automata, but many have evolved mechanisms of motivation and cognition which enable them to use internal as well as external states in determining behaviour (McFarland 1991).

Thus far, McFarland agrees with the situated roboticists of *nouvelle* AI (Chs. 7 and 8). He goes beyond them in using ethological data to suggest a way of analysing the behaviour of any creature. He represents the selective pressures on each species as a cost-function, showing why certain species, with certain behaviours, do or do not flourish within a given niche. The cost to the animal can be measured in terms of the risk of danger, damage, or death—including the additional risks incurred by risk-avoiding behaviour and by changing from one behaviour to another.

McFarland applies his analysis also to robots, including the bomb-disposal robot imagined in a philosopher's thought-experiment on the frame problem (Dennett 1984). The robot's task, in a nutshell, was to retrieve its precious spare battery from a locked room before the time-bomb, also in the room, exploded. This sounds easy enough, until one is informed that the bomb had been placed on the same wheeled wagon as the battery itself. Given that the robot had no explicit warning of this fact, it would need some time to calculate (all) the unintended side-effects of its action of pulling the wagon—which, of course, include the bomb's moving together with the battery.

McFarland points out that, because learning changes the cost–benefit function, bomb-disposal robots able to learn how to increase the availability or accessibility of resources (reaching or defusing the bomb more quickly) would be better adapted to this task-environment than the ones featured in the thought-experiment. One might say this is obvious—and so it is, in qualitative terms. But McFarland can measure how much more adaptive one behaviour is, in certain circumstances, than another.

EXPLANATORY STRATEGIES IN A-LIFE

If A-Life methodologies differ, so too do the explanations favoured by A-Life researchers. Various explanatory strategies are defended in Chapters 7–10. These are relevant not only to the philosophy of A-Life, but to cognitive science as a whole.

One key dispute (prominent in Chs. 7–9) concerns the commitment of classical (and most connectionist) AI and cognitive science to the internal *representation* of *concepts*. Some philosophers have argued that cognitive science must posit symbolic representations, and a combinatorial semantics, like those characteristic of classical AI (Fodor 1975; Fodor and Pylyshyn 1988). Others claim that connectionism, too, allows for (distributed) representations, capable of supporting a combinatorial semantics (A. C. Clark 1989, 1993; Cussins 1990). In Chapters 7–9 the concern is whether A-Life and AI need to posit internal representations at all.

In 'From Robots to Rothko: The Bringing Forth of Worlds' (Ch. 7), Michael Wheeler criticizes the fundamental assumptions of 'Cartesian' cognitive science (orthodox AI and most connectionism). He rejects its commitments to the subject/object distinction and to theories of internal representation. He draws inspiration instead from ecological psychology (Gibson 1979) and from Heidegger and Gadamer, endorsing their stress on the mutual definition of 'subject' and 'environment' and on embodiment and embeddedness as grounds of meaning. And he recommends dynamical systems theory as a non-computational way of thinking about cognition (cf. Beer 1995*a*, 1995*b*; van Gelder 1995; Port and van Gelder 1995; Thelen and Smith 1993).

A-Life and AI are addressed by Wheeler in broadly Heideggerian terms. Heideggerian (and Wittgensteinian) critiques of AI are not new (Dreyfus 1979; Dreyfus and Dreyfus 1986). However, they usually show scant sympathy for the scientific aim of explaining how meaningful behaviour is possible (remarks like 'We just do it' are legion). Wheeler accepts this aim, although he grounds it in a Heideggerian epistemology, and pays attention to the details of the technical literature. Such critiques, moreover, usually accept Heidegger's view that language and culture are essential to world-making. But Wheeler does not. In that sense, his position is non-Heideggerian.

Wheeler argues that ethology and cognitive science should not assume that creatures adapt to objective properties of the world. It is not enough to say that different animals inhabit different worlds in the sense that their sensory-motor capacities enable them to respond only to selected aspects of the objective world (von Uexkull 1957). Nor is it enough to point out, as ecological psychologists do, that the environment offers different 'affordances' to different species, according to their sensory-motor equipment and behavioural repertoire. Rather, these sciences should focus on how situated, subject-engendered, meanings arise from the bodily actions of creatures (animals or animats) embedded in their self-generated world.

However, Wheeler rejects Heidegger's claim that human culture is essential to meaning. For this would imply that languageless animals do not live in meaningful worlds, and that naturalism in the philosophy of mind is false. According to Wheeler's 'hermeneutic naturalism', animals construct their own significant ecological worlds in exploiting the fitness affordances presented to them (cf. Wheeler 1995). Explanations of intentionality based on subject-engendered meanings need not involve language, and can show, in a way continuous with natural science, how a physical system can also be an intentional system.

Non-linguistic meanings are crucial even for some distinctively human phenomenology. Wheeler argues that our experience of the visual (and musical and performing) arts consists largely of non-linguistic meanings that arise from our embeddedness in the world. These meanings are generated by the interaction between the viewer and the work, an aesthetic moment to be characterized in neo-Heideggerian (not 'Cartesian') terms. Significantly, for Wheeler, this characterization finds naturalistic support in (a reformed) cognitive science.

Wheeler's scepticism about representational explanation is not shared by David Kirsh or Andy Clark (Chs. 8 and 9). They both see intelligent behaviour, even in non-human animals, as requiring representations of some sort—or, better, of some sorts. Situated robotics already uses representations in some (non-symbolic) sense; and symbolic, conceptual, representations are needed for full human intelligence.

Chapter 8 is Kirsh's paper, 'Today the Earwig, Tomorrow Man?' Kirsh argues that the ecological approach typical of situated robotics cannot capture those types of perception, learning, and control which require concepts. These intelligent capacities far surpass the powers of insects, which situated robotics takes as its model. (More accurately: which it took as its model until very recently. The leaders of situated research have embarked on a project to build a humanoid robot (Brooks and Stein 1994; Dennett 1994). This will model some aspects of infant development, although developmental epigenesis will be underplayed (Rutkowska 1995). These roboticists intend to follow the ecological path as far as they can, but allow that symbolic representations and internal search may be needed at the final stages.)

Kirsh commends situated robotics for extending the domain of concept-free action, and for showing (as argued also in Ch. 6) that articulated world-models are less crucial than classical cognitive science assumes. He insists, however, that its bottom-up approach has explanatory limitations which only systematic conceptualization can overcome.

Full-blooded concepts enable a creature to recognize perceptual invariance; to reify and combine invariances (referring predicates to names, or drawing inferences); to reidentify individuals over time; to engage in anticipatory self-control; to negotiate between (not just to schedule) potentially conflicting desires; to think counterfactually; to use language to create new abilities; and, by teaching these abilities to others, to make cultural evolution possible. Adult human beings possess all these capacities, chimps most of them, dogs some of them, and new-born babies hardly any. Kirsh does not conclude that non-human animals must possess symbolic representations, or compositional internal notations. Their (non-linguistic) concepts may be implemented in their computational architecture in other ways. But logic and language, and thoughtful human action, require symbolic computation.

In 'Happy Couplings: Emergence and Explanatory Interlock' (Ch. 9) Clark, too, defends representationalism. Criticizing cognitive science's excessive focus on symbolic world-models, he posits a range of partial, egocentric, representations that are constructed (and destroyed) in the course of embodied action. Such representations are used within *nouvelle* AI, and there is neurophysiological evidence for them also.

These representations are crucial to 'interactive' explanation, which stresses the ongoing role of the environment in the control of behaviour. Interactive explanation can often be underwritten by the 'homuncular' variety, which explains the system's behaviour in terms of the actions of its parts (including the mechanisms that respond to environmental cues).

The third explanatory style distinguished by Clark is emergent explanation. Emergence is widely used as a key term by A-Life workers, as indicated for example by the passages (above) on Langton and Kauffman. The concept has a long pedigree in the philosophy of biology, where it is recognized as problematic (the entry for 'emergent evolutionism' in the *Encyclopedia of Philosophy* fills three large pages), but it is rarely explicitly analysed by A-Life

researchers (but see Cariani 1991; Steels 1991). Broadly, it refers to a situation, and/or a type of explanation, where some genuine novelty arises out of a lower level. The difficulty lies in clarifying just what 'novelty' means here.

Clark resists the tendency, common in A-Life, to define emergence in terms of what is unexpected. Something unexpected by most people may be confidently anticipated by a competent mathematician. Nor does Clark define emergence in terms merely of the need for a theoretical vocabulary different from that which describes the lower-level components. For him, emergence involves behaviours (or internal system-properties) that are not caused by any directly controllable part or environmental interaction, but which are side-effects of directly controllable actions and/or interactions.

Because one cannot specify an inner state responsible for the emergent property, this third type of emergence is even more 'anti-reductionist' than the second. But it is not mysterious. It can be explained by showing how it arises (indirectly) from internal and/or interactive phenomena. These may involve small numbers of different parts (requiring componential analysis), or large numbers of similar parts interacting in complex ways. In the latter case, dynamical systems theory may be useful (Port and van Gelder 1995). It is well suited to describe system-environmental coupling and multiply interacting units, and—because it ignores the component structure—it can focus on emergent properties as such.

Clark argues that emergence can be fully understood only if we also consider the system's inner structure. The mechanisms by which emergent properties are (indirectly) generated, *pace* Wheeler (Ch. 7), may often involve homuncular decomposability, and representations too. Theoretical biology and cognitive science must integrate various explanations. These concern adaptive organism-environment coupling; the cooperative activity of the underlying neural components (brain-modules as well as neurones); and the computational (including representational) roles of those components.

Emergence features strongly also in Chapter 10. In his paper 'In Praise of Interactive Emergence, Or Why Explanations Don't Have to Wait for Implementations', Horst Hendriks-Jansen rejects explanations based on componential, functional analysis. Instead, he recommends a form of historical explanation that highlights the emergence of structured behaviours on successive levels.

Drawing on situated robotics, ethology, and developmental psychology, he shows how novel behaviours can emerge during the interactions of a creature with its world. That world is not (objectively) pre-categorized, but is constructed via the creature's own activities—to which, in turn, other creatures may react. Human thought and language are emergent activities, made possible by the 'scaffolding' presented by interactive turn-taking between mother and baby, and by various inborn behaviours which function to attract, and to keep, the adult's attention. These are examples of what Langton calls the predictability gap between genotype and phenotype (Ch. 1), and of what Piaget calls epigenesis (Boden 1995, Preface and ch. 6).

According to Hendriks-Jansen, such species-typical behaviours (discoverable by ethological observation) are the true natural kinds of behavioural science. A *behavioural* science is the most we can hope for, since there is usually no systematic mapping between behaviour and specific internal mechanisms: most behaviour is emergent in Clark's (third) sense. Cost–benefit accounts of behaviour (Ch. 6) are useful abstractions, but do not explain how the observed behaviours arose. Historical theories do explain this, and are testable by situated and evolutionary robotics.

Chapters 9 and 10 indicate that, although A-Life's potential contribution to biology is undeniable, its importance for a truly general cognitive science is more problematic. A-Life could contribute to such a science if there were an explanatory continuity between terrestrial biology and psychology, and between their non-earthbound equivalents—that is, if psychological explanation requires reference to biological constraints (Elton 1994). But this may not be so.

For instance, Clark's discussion of emergence—notwithstanding his call for explanatory interlock—suggests that there may be no close mapping between psychology and neurophysiology, still less any principled mapping to explain intelligent life in general. And Hendriks-Jansen argues that no useful cognitive generalizations can be made at the level of mechanism, and that psychological explanations must be behaviour-based. A-Life (and *nouvelle* AI) differs from classical AI in not focusing on top-down functional decomposition and modularity. In not insisting on a systematic mapping from global behaviour to biological mechanisms (or to vital functions in general), A-Life leaves open the possibility that it may not feature in a general cognitive science.

WHAT IS LIFE?

The concept of life itself is the topic of the remaining papers. Chapters 11–13 address the concept directly, with some consideration of its use in A-Life; Chapters 14 and 15 focus on whether cybernetic entities inhabiting virtual computer-worlds (such as Tierra) are really alive. As these papers show, there is disagreement about what life is, and even about what sort of thing—individual or evolving population—is best regarded as alive. Nevertheless, certain properties are cited repeatedly.

Gareth Matthews (Ch. 11) reminds us that some of these properties were noted long ago by Aristotle. His list included: self-nutrition, growth, decay, reproduction, appetite, sensation or perception, self-motion, and thinking. Some of these are less clear than they might seem. (For instance, mules are sterile, and young animals and some adults cannot reproduce; so perhaps 'reproduction' is satisfied by having been generated by reproduction, not only by being capable of reproduction oneself.) Moreover, there are inconsistencies in Aristotle's account of how these properties relate to the definition of life. Matthews therefore suggests that we define the concept, instead, in terms of Aristotle's notion of powers that tend towards the preservation of the species.

Aristotle's account of life is regarded by some philosophers as an anticipation of modern functionalism, and a praiseworthy attempt to outline a science of life and mind (Nussbaum and Putnam 1992; Wilkes 1992). This interpretation has been disputed, on the grounds that functionalism is a response to Descartes's mind–body problem, which did not arise for Aristotle, and that Aristotelian physics was so fundamentally different from modern science as to be worthless (Burnyeat 1992). Hilary Putnam, the founder of functionalism, remains unconvinced by this argument. He has abandoned his initial definition of functionalism, because of the multiple realizability of mental processes in computational, as well as material, terms (Putnam 1988). Even so, he sees Aristotle's thought about life and mind as similar to functionalism, and as a valuable contribution to the philosophy of the life-sciences. Aristotle offers, he says, 'the fulfilment of Wittgenstein's desire to have a [non-reductive] "natural history of man"' (Nussbaum and Putnam 1992: 56). Such readings interpret Aristotle in a way consonant with the assumptions of A-Life in general, and with Langton's stress on organization in particular.

Aristotle saw no metaphysical chasm between life and mind (*psyche*)—rationality (*nous*) perhaps excepted—or between biology and psychology. He saw life and mind as involving progressively powerful forms of organization, the capacities possessed by a lower form being possessed also by all higher forms. On that view, it should be no surprise if a philosophy (such as functionalism) originally developed to deal with mind can be generalized to life.

Descartes, by contrast, saw mental events (and *a fortiori* rational thought) as enjoyed only by humans. There can be no science of animal psychology, since animals are merely complicated mechanical systems. Humans have minds not because psychological properties emerge from the complex organization of a living material base, but because an extra ingredient (mental substance) is added by God. Granted, the lamb runs on seeing the wolf. But what Aristotle would have explained in terms of perceptive powers, Descartes explained in terms of the movements of the animal spirits within the lamb's body. These movements are described, in principle, by physics—which can also explain the living functions of animals and plants.

For Aristotle, too, life is grounded in physics. But unlike Descartes he did not ask 'How can material things possibly perceive?', or even 'How can matter possibly be alive?', since he assumed that physics naturally allows the emergence of increasingly powerful forms of organization. The assumption that matter somehow has a potential for developing life is shared by A-Life. Research on auto-catalytic networks and artificial metabolisms could be viewed as attempts to define the general form of this self-organizing potential, and to identify unfamiliar material carriers of it. Kauffman's work, for instance, suggests that life (and mind) is a possibility inherent in matter, whose realization is much more probable than is commonly assumed (Ch. 4).

Philosophies of life, and A-Life methodologies also, differ on various dimensions identified by Peter Godfrey-Smith in Chapter 12: 'Spencer and Dewey on Life and Mind'. He distinguishes internalism and externalism; asymmetrical and symmetrical externalism; and weak and strong versions of the ontological continuity of life and mind.

Internalist approaches see life as a process of autonomous self-organization, wherein internal constraints govern the history and interactions of the constituent units of the system. Externalists, by contrast, explain the system's internal structure primarily as a result of its adaptive interactions with the environment. Both Spencer and Dewey favoured externalist views of life, although Spencer gave an internalist account of the evolution of the universe prior to the origin of life. Within A-Life, work on cellular automata and Kauffman's discussions of self-organizing networks are internalist. Research on evolutionary robotics, and most other work on animats, is externalist.

The asymmetric externalist (such as Spencer) emphasizes the organism's adaptive reactions to the environment. The ways in which the organism, in its turn, changes the environment are downplayed, though not necessarily denied. In symmetric externalism, this unceasing feedback is highlighted, and the self-adaptive organism is shown to play an active role in shaping its ecological world. Applied to A-Life, asymmetry characterizes situated robotics, most animat research, and many uses of GAs. Symmetry is stressed in research on co-evolution (including co-evolutionary robotics), and in A-Life work that sees the organism and environment as closely coupled dynamical systems, between which no principled distinction—and no epistemological 'gap'—exists. Arguments against 'Cartesian' cognitive science (Ch. 7) are reminiscent of Dewey's (symmetric) criticism of the separation of organism and environment.

Like Aristotle (but unlike Descartes), both Spencer and Dewey were continuity theorists. That is, they saw human minds as ontologically continuous with lower-level vital capacities, from which they have emerged. But there are weak and strong interpretations of continuity. The weak version asserts that mind implies life, that mind can arise only in living things. On this view, AI must be preceded by A-Life. Strong continuity asserts that minds are 'literally life-like', in that life and mind share basic organizational principles. On this view, AI is a sub-class of A-Life. (Godfrey-Smith also defines 'methodological' continuity, the view that a scientific understanding of mind is in practice achievable only if mind—perception, cognition, motivation—is studied in the context of the entire organism. *Nouvelle* AI, and much work in A-Life, favours methodological continuity.)

Whether Aristotle was a weak or a strong continuity theorist is disputed (Ch. 11). As for Spencer, he accepted both strong and methodological continuity. He saw intelligence as emerging from life when serial processing replaces parallel processing. And, because he defined life in (functionalist) terms of self-sustaining adaptation to the environment, Godfrey-Smith suggests that he would have described Ray's Tierra 'organisms' as alive, despite their lack of biochemical metabolism. Dewey's position is less clear. He certainly accepted weak continuity. But he favoured strong continuity only for certain aspects of mentality—language specifically excluded. Godfrey-Smith suggests that, for Dewey, A-Life could lead to genuine AI only if it studied societies of intercommunicating creatures. Current A-Life studies of (for example) the use of pheromones in ant-colonies would not qualify, since for Dewey communication is symbolic.

Mark Bedau, in 'The Nature of Life' (Ch. 13), compares various definitions of life and argues that the most satisfactory is 'supple adaptation'. This is a near-synonym for evolution, but it makes explicit the flexibility with which living things adapt to a host of unpredictable environmental contingencies which continually alter the selective criteria for fitness. And, crucially, it covers both the puzzling unity and the astonishing diversity of life-forms.

Bedau uses his definition to conceive of life as a natural kind, not an arbitrary list or cluster of properties. He admits that his criterion applies to a number of systems we would not normally regard as alive, such as human cultures and economic markets. But he does not present it as a philosophical analysis of our intuitive, pre-theoretic concept of life. He justifies it instead on scientific grounds, acknowledging that it sometimes fails to map on to our everyday concept (as the chemist's 'water' does, too).

One of the ways in which Bedau's concept of life differs from the everyday one is in its logical category. He argues that we should conceive of an individual's life in relation to the vitality of the global system, or evolving population, which makes the individual organism possible. This view is implicit in his criterion of supple adaptation. If one takes the paradigm case of a 'living thing' to be a single organism, then Bedau's definition of life is problematic. A single organism, whether microbe or mammal, is rarely flexible in adapting to new circumstances even if 'adapting' is interpreted as 'learning', and is never adaptable in the sense of being able to evolve. Only species, or lineages of organisms, can do that. Bedau ascribes life first of all to evolving populations, and only secondarily to individual living things.

Bedau's definition may be seen as problematic in two other ways. In building evolution into his definition of life, he has to admit that an evolved population which had stabilized in equilibrium (there being no more environmental changes, and no more co-evolutionary pressure from other species) would not fully satisfy his criterion. He does not take a strongly essentialist position: 'Well then, the population simply wouldn't be alive.' Rather, he points out that this imaginary population would have originated through an evolutionary process. He adds that such a situation is empirically highly improbable, because the evolutionary stability will in practice always be temporary, so the example has no real purchase for biological science.

A second problem for Bedau, and for any biologist who (like Maynard Smith) defines life in terms of evolution, is the implication that creationism is incoherent. Creationism claims that each living species was specially created by God. But such species would not be alive at all, according to an evolutionary definition: unlike the species in Bedau's evolution-in-equilibrium scenario, they do not even have evolutionary descent somewhere in their history. Biologists who oppose creationist 'theory' (still taught in biology classes in some schools) normally argue that it offers no principled explanation of what Bedau calls the unified diversity of life, whereas neo-Darwinism does. Someone genuinely trying to decide whether creationist or neo-Darwinian schoolbooks offer a better explanation of life might be persuaded by such arguments, but not by being told that life involves evolution *by definition*. To argue that 'creation biology' is false, or explanatorily inferior to scientific biology, one must treat evolution as a universal empirical characteristic of life, not as an *a priori* criterion of it.

A-Life interests Bedau primarily because it can clarify the notion of supple adaptation. In doing so, it also clarifies the sense in which evolution is teleological, for A-Life offers quantitative measures of functionality—some developed by Bedau himself (see also Cliff and Miller forthcoming). The well-known example of the spandrels of San Marco (Gould and Lewontin 1979) shows how difficult it is to decide which traits are functional and which are merely side-effects. It is even more difficult to measure and compare the contribution to 'fitness' of different functional traits. Bedau sees A-Life as able to give precise answers to these questions, allowing that the criteria for selection themselves change as the system evolves. He applies his functionality-metric to a number of examples, including the co-evolving Tierra system described in Chapter 3. Not least, in the present context, he offers a measure of 'vitality' (in terms of the rate at which new significant adaptations are arising and persisting) which might help us to decide to what extent a given system involves life.

As is evident from Chapters 1, 3, and 11–13, the characteristics repeatedly mentioned in discussions of the nature of life include self-organization, autonomy, responsiveness, reproduction, evolution, and metabolism. Arguably, each of these

(except one) is definable in informational terms, and each (except one) has already been exhibited to some degree in functioning A-Life systems. The exception, apparently, is metabolism. Systems like Tierra show real self-replication, and real co-evolution. But whether the things that are self-replicating and co-evolving are really alive depends, one might argue, not just on their computational properties but also on their physical embodiment and metabolism.

Is metabolism denied to A-Life systems? Certainly, we cannot confine the concept to energy-exchanges involving terrestrial proteins, phosphates, and oxygen. One might argue that metabolism necessarily involves both anabolism and catabolism: the breaking down of some chemicals and the synthesis of others. In that case, the electrical energy-exchanges within a computer would not qualify; nor would the energy-use of a manufactured robot. But what if metabolism means (more broadly) energy-use resulting in self-repair, and in the other vital characteristics listed above? A future Tierra might evolve a way of recognizing and repairing transcription-errors arising because of unpredictable physical events in the hardware (compare the body's recognition and repair of a tear in a blood-vessel or of faulty copying of DNA). The computer's energy would be required for this, as it is for reproduction and evolution.

Perhaps such energy-use cannot reasonably be counted as metabolism, since it is not the case that each Tierran 'organism' has its own individuated energy-budget. If this constraint could be added, the likeness to biological metabolism would be greater—but there would still be no analogy of anabolism and catabolism (or of the complex chemical life-cycle resulting from them).

The importance of physical embodiment and energy-exchange enters the discussion in Chapters 14 and 15, which ask whether A-Life can ever be real life. Could Langton or Ray succeed in their aim of synthesizing novel life-forms? Recalling Searle's (1980) discussion of AI, Elliott Sober addresses this question in Chapter 14: 'Learning from Functionalism—Prospects for Strong Artificial Life'.

Sober's rejection of strong A-Life depends not on any definition of life but on his analysis of specific biological properties of terrestrial living things. Much as intentionality requires a causal world–mind relationship, not captured in abstract functionalist terms, so some biological properties relate an organism to something outside itself. It follows that computers—or computational processes running on computers, such as computer viruses—cannot really have these properties (although they may replicate some aspects of them) unless they are related to their environment in the relevant ways (cf. Harnad 1994*a*, 1994*b*).

Examples cited by Sober include reproduction, digestion, and predation. He allows that reproduction, however it may be physically realized, always involves the formation and manipulation of representations (DNA instructions, for instance). Reproduction, as von Neumann claimed, is essentially computational.

But digestion, also realized in multiple ways, is not. Whereas an A-Life model might really reproduce, it could not digest—or rather, it could not digest *in virtue of its being a computational system* (though some future animat might employ a novel digestive biochemistry discovered by A-Life). Again, since predation involves both hunting and eating, true predation requires that the predator really consume the energy of the prey. In other words, metabolism (and associated behaviours, such as predation) necessarily involves a physically embodied organism interacting with a physical environment. Computer-code creatures within virtual environments do not qualify.

Sober's close assimilation of A-Life to AI has been challenged on the grounds that two familiar objections to strong AI cannot counter strong A-Life (Keeley 1994). Eliminativist arguments against folk-psychology are prominent in AI and cognitive science. Some eliminativists see everyday psychological terms (belief, desire, hope, regret) as instrumental categories, useful as theoretical shorthand but not denoting real entities—or, at best, as categories denoting *abstracta* (comparable to centres of gravity), not *concreta* (Dennett 1987). Others foresee their eventual replacement by neuroscientific categories as yet unknown (Churchland 1989). Such arguments are widely regarded with scepticism; few philosophers are prepared to countenance the disappearance of mentalistic vocabulary in favour of some future neuroscience. But eliminativism is not common in A-Life, *nouvelle* AI excepted (Brooks 1991; Smithers 1992), because 'folk biology' is broadly consonant with scientific biology—and it avoids immaterialist notions such as Bergson's *élan vital*. Accordingly, anti-eliminativist arguments have little purchase against A-Life, except in rebutting claims that animals (and animats) need no representations of any kind (see Chs. 8 and 9).

The second difference between strong AI and strong A-Life concerns consciousness. Whereas consciousness prompts many attacks on strong AI (and functionalism), it is hardly mentioned in discussions of A-Life—or in biology itself. Philosophers who believe there is something it is like to be a bat (Nagel 1974) might object to McFarland's vision of animals as cost-based robots, which (unlike Wheeler's hermeneutic naturalism outlined in Ch. 7) leaves no room for subjectivity. But consciousness is irrelevant to self-organization, self-replication, or evolution. Only someone who believed that all living things—plants and animals alike—have some sort of mind, or awareness, would dismiss A-Life as a whole because it (almost always) ignores consciousness.

Strong A-Life is discussed also in Chapter 15: 'Simulations, Realizations, and Theories of Life'. Howard Pattee sees evolution as essential to life, and as requiring symbolic genotypes, material phenotypes, and selective environments. Since realizations are material models that actually implement the relevant functions, strong A-Life requires the evolution of novel material phenotypes, performing novel functions, in input–output relations with their environment.

Animats produced by evolutionary robotics might qualify. But what about digital phenotypes in purely formal environments, such as Ray's *Digital Reserve*? Formal environments, says Pattee, can support real life only if they enable the emergence of fundamental novelty in evolution. His analysis of emergence requires that qualitatively new phenotypic structures and functions arise in unpredictable ways. An open-ended physical environment interacts unpredictably with both genotype and phenotype. These (epigenetic) physical interactions, and the selective forces acting on the organism, are not describable by the genotype (considered as symbolic information), and lead to the evolution of features not wholly prefigured in it.

Among the most important emergent novelties, for Pattee, are new forms of measurement (cf. Cariani 1991; Rosen 1991). He defines measurement as a physical interaction resulting in a classification—a definition that covers any observation of, or information-bearing response to, the world. The classification may be based in some novel sensory mechanism, and may result in an enduring informational record. But it may merely be a mapping from a physical pattern to a specific motor action, or even biochemical activity (enzyme synthesis, for instance). The physical basis of the classification is not an inevitable consequence of physico-chemical laws, but is arbitrary—as in the genetic code, for instance. Since physical interaction is essential, and material systems cannot be realized by formal systems (which deal only with syntax, not semantics), novel types of measurement could not emerge by purely computational means.

Pattee adds that the notion of measurement (observation) is problematic. It poses the epistemological problem of the origin of meaning, or information, from matter—and its corollary, the puzzle of how information (e.g. the genotype) can constrain physical processes (the phenotype). Pattee locates the difficulty in the fact that measurement is irreversible whereas physical laws are in principle reversible—yet all measurements must be effected by physical means. Quantum physics, instead of regarding measurement as an extra-theoretical determination of initial conditions, places it at the heart of physical theory. This leads to fundamental difficulties in distinguishing physical from informational processes. Some quantum theorists appeal to consciousness in making this distinction, while others demand only that some record be irreversibly made; however, *consciousness* and *record* are deeply problematic concepts. In short, some central concepts of biology (and psychology) are fundamentally unclear. Pattee allows that (weak) A-Life might help clarify them, but insists that (strong) A-Life could not realize them.

Elsewhere Pattee (1995) argues that A-Life models, considered as ideal formal systems, cannot be perfectly realized in physical machines but only approximately simulated by them. The requirements for a formal system include some—such as absolute determinism in reading and writing—which are not physically realizable. It follows that no purely computational A-Life (or AI) system can be faithfully translated into physical form, whether a computer implementation of some virtual environment or a biochemical system in the real world.

In general, the genotype—considered purely as a symbolic description—cannot be perfectly realized in the phenotype. (For example, genes specifying the linear sequence of amino-acids in a protein molecule do not control how the constructed molecule will fold: that is determined by the molecule's intrinsic physical properties.) It is this which allows for the possibility of 'open-ended' evolution, with the emergence of new forms of vital (meaningful) function. But the philosophical problem of deriving meaning from matter remains. Pattee calls it the 'intrinsic epistemic cut between the genotype and phenotype, i.e. between description and construction', and argues that we do not adequately understand what *implementing a description* means.

A-LIFE AND PUBLIC LIFE

Many chapters in this volume involve attempts to define necessary, or even necessary and sufficient, conditions of life. Some philosophers will have scant sympathy for such essentialist projects, arguing that what counts as alive is a question of how the term is actually used. But usage can change, not least under the influence of science: *evolution* has already been added to biologists' criteria of life. Technology, too, can affect it. Turing's (1950) prediction that by the millennium computers will routinely be described as thinking looks likely to be confirmed, so that—despite the philosophical inadequacy of the Turing Test—people are tacitly predisposed towards philosophical positions which they might otherwise reject. Will A-Life have a similar effect?

Children's understanding of life is already being challenged by widely available A-Life technology, such as home-computer versions of the *biomorph* (Dawkins 1986, appendix; Dawkins 1991) and Tierra programs. Youngsters of the 1970s mentioned life as one of two features distinguishing people from computers, the other being feeling or emotion (Turkle 1984). They conceded that computers are intelligent, but not that they are alive. This rejection of the continuity thesis involved few attempts to analyse 'life', because the concept was not challenged by traditional computing. With Tierra in the living-room, however, this concept becomes a focus of discussion (Turkle 1995, ch. 6).

Children of the 1990s mention several characteristics of life, the most important being autonomous movement. This feature was emphasized also in Piaget's (1929) classic study of the child's concepts of life and mind. Children faced with A-Life, however, may distinguish between movement on the VDU-screen

and movement of physical bodies. Only the latter, they may say, counts as evidence of life; and even bodily movement may be discounted, if the bodies are not natural but artefactual. In addition, these children think of life in terms of reproduction and evolution, both of which they can observe on their computer screens.

But, unsurprisingly, they offer no agreed list of necessary and sufficient conditions. Children (and adults) disagree—and vacillate—over whether A-Life creatures are really alive as we are, alive merely as ants are, sort of alive, or not alive (but nevertheless in control). Sometimes, these distinctions are brushed aside with the remark: 'It's just a machine.' (The notion of *machine* has itself altered since Piaget's day, and even since the early days of home-computing around 1980: in the 1990s, people acknowledge the surprising, apparently magical powers of many A-Life systems.) Very few people are willing to assert (with Ray) that Tierran creatures are just as alive as biological organisms. In general, the distinction between humans and machines is protected—if only by mutually inconsistent definitions, used in different circumstances. The papers in this volume show that whether philosophers can justify this everyday distinction in rigorous terms is controversial.

Besides prompting changes in everyday notions of what it is to be alive, A-Life may also affect people's attitudes to technology in general. These attitudes have been described as falling into two broad types: 'Faustian' and 'Magian' (S. Clark, in press). The first, nurtured by the scientific and industrial revolutions, sees technology as under our control. Science renders the world intelligible, and technology reliably exploits this intelligibility (and deterministic predictability) to help us achieve our goals. The second sees technology as shot through with dangerous unpredictability, and its scientific underpinnings as unintelligible except to a tiny élite. Not scientist as boffin, but scientist as magician—whose arcane technology threatens us in largely hidden ways.

One example of the way in which public confidence in technology as control has already been eroded is the ecological movement. Concern about the damage done by technology to the global environment ranges from the near-mystical Gaia hypothesis (Lovelock 1979) to the sober realization that 'minor' technological side-effects (CFCs, for instance) can upset delicate and far-reaching equilibria—with potentially dire results. The ecological movement constantly emphasizes the unexpected harm that can be caused by apparently benign technologies, whose wider effects we are only beginning to glimpse.

A-Life can be expected to arouse similar suspicions, for it is widely unintelligible and largely occult. The workings of an 'engine' whose entire function consists of electronic dances across computer circuitry are invisible. Add the emergence of new order and the evolutionary randomness typical of A-Life, and the aura of magic intensifies (Kelly 1994). Interviews with adults and children show that many people experience A-Life phenomena as shocking, magical, and godlike (Turkle 1995, ch. 6). Even A-Life professionals will not have a detailed understanding of programs or robots evolved over many generations in a complex task-environment. Such systems will function largely as black boxes. Ray's vision of the *Digital Reserve*, despite his promise that its denizens will not escape to infect computers across the Internet, is uncomfortably reminiscent of the ill-understood and ultimately malignant magic of the sorcerer's apprentice.

Images of technology are deeply rooted in our culture, and feed back into it. A-Life will play its role in this cultural process. If even professionals cannot fully understand it, laymen may simply stop trying. The prominence of Magian technologies could thus encourage irrationalism, even anti-rationalism. More helpfully, such technologies could prevent us from assuming (with Faust) that we can be clear about all our goals, and can reasonably hope to realize just what we want—and no more. That Enlightenment ambition still survives, to a considerable degree. A-Life, despite what many see as its overweening *hubris*, might help to make our ambitions less vainglorious.

REFERENCES

Ballard, D. H. (1991), 'Animate Vision', *Artificial Intelligence*, 48: 57–86.

Beer, R. (1995*a*), 'Computational and Dynamical Languages for Autonomous Agents', in R. Port and T. van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition* (Cambridge, Mass.: MIT Press).

—— (1995*b*), 'A Dynamical Systems Perspective on Environment–Agent Interactions', *Artificial Intelligence*, 72: 173–215.

Bersini, H., and Detours, V. (1994), 'Asynchrony Induces Stability on Cellular Automata Based Models', in Brooks and Maes (1994), 382–7.

Boden, M. A. (1980), 'The Case for a Cognitive Biology', *Proceedings of the Aristotelian Society*, Supp. vol. liv, 25–49. (For an extended version, see M. A. Boden, *Minds and Mechanisms: Philosophical Psychology and Computational Models* (Ithaca, NY: Cornell University Press, 1981), 89–112.)

—— (1987), *Artificial Intelligence and Natural Man*, 2nd edn., expanded (London: MIT Press).

—— (1990) (ed.), *The Philosophy of Artificial Intelligence* (Oxford Readings in Philosophy; Oxford: Oxford University Press).

—— (1995), *Piaget*, 2nd edn., expanded (Fontana Modern Masters Series; London: HarperCollins).

Braitenberg, V. (1984), *Vehicles: Essays in Synthetic Psychology* (Cambridge, Mass.: MIT Press).

Brooks, R. A. (1991), 'Intelligence Without Representation', *Artificial Intelligence*, 47: 139–59.

—— and Maes, P. (1994) (eds.), *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems* (Cambridge, Mass.: MIT Press).

—— and Stein, L. A. (1994), 'Building Brains for Bodies', *Autonomous Robots*, 1: 7–25.

Burks, A. W. (1970) (ed.), *Essays on Cellular Automata* (Champaign-Urbana: University of Illinois Press).

Burnyeat, M. F. (1992), 'Is an Aristotelian Philosophy of Mind Still Credible?', in Nussbaum and Rorty (1992), 15–26.

Cariani, P. (1991), 'Emergence and Artificial Life', in Langton *et al.* (1991), 775–97.

Churchland, P. M. (1989), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (Cambridge, Mass.: MIT Press).

Clark, A. C. (1989), *Microcognition: Philosophy, Cognitive Science, and Parallel Distributed Processing* (Cambridge, Mass.: MIT Press).

—— (1993), *Associative Engines: Connectionism, Concepts, and Representational Change* (Cambridge, Mass.: MIT Press).

Clark, S. R. L. (in press), 'Tools, Machines, and Marvels', in R. Fellowes (ed.), *Philosophy and Technology* (Cambridge: Cambridge University Press).

Cliff, D. (1991), 'The Computational Hoverfly: A Study in Computational Neuroethology', in Meyer and Wilson (1991), 87–96.

—— and Miller, G. F. (forthcoming), 'Co-Evolution of Pursuit and Evasion II: Simulation Methods and Results', submitted to *Adaptive Behavior*.

—— Harvey, I., and Husbands, P. (1993), 'Explorations in Evolutionary Robotics', *Adaptive Behavior*, 2: 71–108.

—— Husbands, P., Meyer, J.-A., and Wilson, S. (1994) (eds.), *Animals to Animats 3: Proceedings of the Third International Conference on the Simulation of Adaptive Behavior* (Cambridge, Mass.: MIT Press).

Cussins, A. (1990), 'The Connectionist Construction of Concepts', in Boden (1990), 368–440.

Dawkins, R. (1976), *The Selfish Gene* (Oxford: Oxford University Press).

—— (1986), *The Blind Watchmaker* (Harlow: Longman).

—— (1989), 'The Evolution of Evolvability', in C. G. Langton (ed.), *Artificial Life* (SFI Studies in the Sciences of Complexity, vi: Proceedings of the Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems held September 1987, Los Alamos; Redwood City, Calif.: Addison-Wesley), 201–20.

—— (1991), *Blind Watchmaker: The Program of the Book* (PO Box 59, Leamington Spa: SPA).

De Bourcier, P., and Wheeler, M. (1995), 'Aggressive Signaling Meets Adaptive Receiving: Further Experiments in Synthetic Behavioural Ecology', in Moran *et al.* (1995), 762–73.

Dennett, D. C. (1984), 'Cognitive Wheels: The Frame Problem of AI', in C. Hookway (ed.), *Minds, Machines, and Evolution: Philosophical Studies* (Cambridge: Cambridge University Press), 129–52. (Repr. in Boden (1990), 147–70.)

—— (1987), *The Intentional Stance* (Cambridge, Mass.: MIT Press).

—— (1994), 'The Practical Requirements for Making a Conscious Robot', *Philosophical Transactions of the Royal Society, Series A*, 349: 133–46.

Depew, D. J., and Weber, B. H. (1994), *Darwinism Evolving: Systems Dynamics and the Genealogy of Natural Selection* (Cambridge, Mass.: MIT Press).

Dreyfus, H. L. (1979), *What Computers Can't Do: The Limits of Artificial Intelligence*, 2nd edn. (New York: Harper and Row).

—— and Dreyfus, S. E. (1986), *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer* (New York: Free Press, Macmillan).

Eldredge, N., and Gould, S. J. (1972), 'Punctuated Equilibria: An Alternative to Phyletic Gradualism', in T. J. M. Schopf (ed.), *Models in Paleobiology* (San Francisco: Freeman), 82–115.

Elton, M. (1994), 'What Are You Talking About?', *AISB Quarterly*, No. 87: 47–54. (Special Theme: AI and Artificial Life, ed. D. Cliff.)

Feigenbaum, E. A., and Feldman, J. (1963) (eds.), *Computers and Thought* (New York: McGraw-Hill).

Fodor, J. A. (1975), *The Language of Thought* (Hassocks, Sussex: Harvester Press).

—— and Pylyshyn, Z. W. (1988), 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, 28: 3–71.

Frazer, J. (1995), *An Evolutionary Architecture* (London: Architectural Association).

Gardner, M. (1970), 'The Fantastic Combinations of John Conway's New Solitaire Game "Life"', *Scientific American*, 223/4: 120–3.

Gibson, J. J. (1979), *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin).

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning* (Reading, Mass.: Addison-Wesley).

Gould, S. J., and Lewontin, R. C. (1979), 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme', *Proceedings of the Royal Society, B*, 205: 581–98.

Harnad, S. (1994*a*), 'Artificial Life: Synthetic vs. Virtual', in Langton (1994), 539–42.

—— (1994*b*), 'Levels of Functional Equivalence in Reverse Bioengineering', *Artificial Life*, 1: 293–301.

Harvey, I. (1992), 'Species Adaptation Genetic Algorithms: A Basis for a Continuing SAGA', in Varela and Bourgine (1992), 346–54.

Hillis, W. D. (1991), 'Co-Evolving Parasites Improve Simulated Evolution as an Optimization Procedure', in Langton *et al.* (1991), 313–24.

Holland, J. H. (1975), *Adapation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence* (Ann Arbor: University of Michigan Press). (Reissued by MIT Press, 1992.)

—— Holyoak, K. J., Nisbett, R. E., and Thagard, P. R. (1986) (eds.), *Induction: Processes of Inference, Learning, and Discovery* (Cambridge, Mass.: MIT Press).

Husbands, P., Harvey, I., and Cliff, D. (1995), 'Circle in the Round: State Space Attractors for Evolved Sighted Robots', *Journal of Robotics and Autonomous Systems*, 15: 83–106. (Special Issue on The Biology and Technology of Intelligent Autonomous Agents, ed. L. Steels.)

Ingerson, T. E., and Buvel, R. L. (1984), 'Structure in Asynchronous Cellular Automata', *Physica D*, 10: 59–68.

Kauffman, S. A. (1993), *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford: Oxford University Press).

—— and Johnson, S. (1991), 'Co-Evolution to the Edge of Chaos: Coupled Fitness Landscapes, Poised States, and Co-Evolutionary Avalanches', in Langton *et al.* (1991), 325–69.

Keeley, B. L. (1994), 'Against the Global Replacement: On the Application of the Philosophy of Artificial Intelligence to Artificial Life', in Langton (1994), 569–87.

Kelly, K. (1994), *Out of Control: The Rise of Neo-Biological Civilization* (Reading, Mass.: Addison-Wesley).

Langton, C. G. (1986), 'Studying Artificial Life with Cellular Automata', *Physica D*, 10: 120–49.

—— (1991), 'Life at the Edge of Chaos', in Langton *et al.* (1991), 41–91.

—— (1994) (ed.), *Artificial Life III* (Santa Fe Institute Studies in the Sciences of Complexity, Proceedings, 17; Redwood City, Calif.: Addison-Wesley).

—— (1995) (ed.), *Artificial Life: An Overview* (Cambridge, Mass.: MIT Press).

—— Taylor, C., Farmer, J. D., and Rasmussen, S. (1991) (eds.), *Artificial Life II* (Santa Fe Institute Studies in the Sciences of Complexity, 10; Redwood City, Calif.: Addison-Wesley).

Levy, S. (1992), *Artificial Life: The Quest for a New Creation* (New York: Random House).

Lovelock, J. (1979), *Gaia: A New Look at Life on Earth* (Oxford: Oxford University Press).

McCulloch, W. S., and Pitts, W. (1943), 'A Logical Calculus of the Ideas Immanent in Nervous Activity', *Bulletin of Mathematical Biophysics*, 5: 115–33. (Repr. in Boden (1990), 22–39.)

McDermott, D. V. (1976), 'Artificial Intelligence Meets Natural Stupidity', *SIGART Newsletter of the ACM*, No. 57, April. (Repr. in J. Haugeland (ed.), *Mind Design: Philosophy, Psychology, Artificial Intelligence* (Cambridge, Mass.: MIT Press), 143–60.)

McFarland, D. (1991), 'Defining Motivation and Cognition in Animals', *International Studies in the Philosophy of Science*, 5: 153–70.

Meyer, J.-A., and Guillot, A. (1991), 'Simulation of Adaptive Behavior in Animats: Review and Prospects', in id. and Wilson (1991), 2–14.

—— —— (1994), 'From SAB90 to SAB94: Four Years of Animat Research', in Cliff *et al.* (1994), 2–11.

—— and Wilson, S. W. (1991) (eds.), *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (Cambridge, Mass.: MIT Press).

Miller, G. F. (1995), *Artificial Life as Theoretical Biology: How to Do Real Science as Computer Simulation* (University of Sussex, School of Cognitive and Computing Sciences Research Paper).

—— and Cliff, D. (forthcoming), 'Co-Evolution of Pursuit and Evasion I: Biological and Game-Theoretic Foundations', submitted to *Adaptive Behavior*.

Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories* (Cambridge, Mass.: MIT Press).

Mingers, J. (1989), 'An Introduction to Autopoiesis—Implications and Applications', *Systems Practice*, 2: 159–80.

Moran, F., Moreno, A., Merelo, J. J., and Chacon, P. (1995) (eds.), *Advances in Artificial Life: Proceedings of the Third European Conference on Artificial Life* (Granada, 1995) (Berlin: Springer).

Nagel, T. (1974), 'What is it Like to be a Bat?', *Philosophical Review*, 83: 435–50.

Nussbaum, M. C., and Putnam, H. (1992), 'Changing Aristotle's Mind', in Nussbaum and Rorty (1992), 27–56.

—— and Rorty, A. (1992) (eds.), *Essays on Aristotle's 'De Anima'* (Oxford: Clarendon Press).

Pattee, H. H. (1995), 'Artificial Life Needs a Real Epistemology', in Moran *et al.* (1995), 23–38.

Piaget, J. (1929), *The Child's Conception of the World* (London: Routledge, Kegan Paul).

Port, R., and van Gelder, T. (1995) (eds.), *Mind as Motion: Dynamics, Behavior, and Cognition* (Cambridge, Mass.: MIT Press).

Prusinkiewicz, P. (1994), 'Visual Models of Morphogenesis', *Artificial Life*, 1: 61–74.

Putnam, H. (1960), 'Minds and Machines', in S. Hook (ed.), *Dimensions of Mind* (New York: Collier), 148–79.

—— (1967), 'The Nature of Mental States', in id., *Mind, Language, and Reality: Philosophical Papers*, ii (Cambridge: Cambridge University Press, 1975), 429–40. (First published as 'Psychological Predicates', in W. H. Capitan and D. D. Merrill (eds.), *Art, Mind, and Religion* (Pittsburgh: University of Pittsburgh Press, 1967).)

Putnam, H. (1988), *Representation and Reality* (Cambridge, Mass.: MIT Press).
Ray, T. S. (1994), 'An Evolutionary Approach to Synthetic Biology: Zen and the Art of Creating Life', *Artificial Life*, 1: 179–210.
Reynolds, C. W. (1987), 'Flocks, Herds, and Schools: A Distributed Behavioral Model', *Computer Graphics*, 21: 25–34.
—— (1994*a*), 'An Evolved, Vision-Based Model of Obstacle Avoidance Behavior', in Langton (1994), 327–46.
—— (1994*b*), 'Competition, Coevolution, and the Game of Tag', in Brooks and Maes (1994), 59–69.
Rosen, R. (1991), *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life* (New York: Columbia University Press).
Rutkowska, J. (1995), 'Can Development be Designed? What we may Learn from the Cog Project', in Moran *et al.* (1995), 383–95.
Searle, J. R. (1980), 'Minds, Brains, and Programs', repr. in Boden (1990), 67–88.
Sims, K. (1991), 'Artificial Evolution for Computer Graphics', *Computer Graphics*, 25: 319–28.
—— (1994), 'Evolving 3D Morphology and Behavior by Competition', *Artificial Life*, 1: 353–72.
Smithers, T. (1992), 'Taking Eliminative Materialism Seriously: A Methodology for Autonomous Systems Research', in Varela and Bourgine (1992), 31–40.
Spafford, E. H. (1991), 'Computer Viruses—A Form of Artificial Life?', in Langton *et al.* (1991), 727–45.
—— (1994), 'Computer Viruses as Artificial Life', *Artificial Life*, 1: 249–65.
Steels, L. (1991), 'Towards a Theory of Emergent Functionality', in Meyer and Wilson (1991), 451–61.
Thelen, E., and Smith, L. B. (1993), *A Dynamic Systems Approach to the Development of Cognition and Action* (Cambridge, Mass.: MIT Press).
Thompson, A. (1995), 'Evolving Electronic Robot Controllers that Exploit Hardware Resources', in Moran *et al.* (1995), 641–57.
Todd, S., and Latham, W. (1992), *Evolutionary Art and Computers* (London: Academic Press).
Turing, A. M. (1936), 'On Computable Numbers, with an Application to the *Entscheidungsproblem*', *Proc. London Math. Soc.*, 42: 230–65.
—— (1950), 'Computing Machinery and Intelligence', *Mind*, 59: 433–60.
—— (1952), 'The Chemical Basis of Morphogenesis', *Philosophical Transactions of the Royal Society (London), Series B*, 237: 37–72.
Turk, G. (1991), 'Generating Textures on Arbitrary Surfaces using Reaction-Diffusion', *Computer Graphics*, 25: 289–98.
Turkle, S. (1984), *The Second Self: Computers and the Human Spirit* (New York: Simon and Schuster).
—— (1995), *Life on the Screen: Identity in the Age of the Internet* (New York: Simon and Schuster).
Uexkull, J. von (1957), 'A Stroll through the Worlds of Animals and Men', in C. H. Schiller (ed.), *Instinctive Behavior: The Development of a Modern Concept* (New York: International Universities Press), 5–82.
van Gelder, T. (1995), 'What Might Cognition Be if not Computation?', *Journal of Philosophy*, 92: 345–81.
Varela, F. J., and Bourgine, P. (1992) (eds.), *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Cambridge, Mass.: MIT Press).

Webb, B. (1994), 'Robotic Experiments in Cricket Phonotaxia', in Cliff *et al.* (1994), 45–54.
Wheeler, M. (1995), 'Escaping from the Cartesian Mind-Set: Heidegger and Artificial Life', in Moran *et al.* (1995), 65–76.
Wilkes, K. V. (1992), '*Psuche* versus the Mind', in Nussbaum and Rorty (1992), 109–27.
Wilson, S. W. (1991), 'The Animat Path to AI', in Meyer and Wilson (1991), 15–21.
Wolfram, S. (1984), 'Computer Software in Science and Mathematics', *Scientific American*, Sept., 188–203.