In F. Varela and P. Bourgine, eds., Towards A Practice of Autonomous Systems
(Cambridge, MA: Bradford Books/MIT Press, 1992), pp. 494-503.

Philosophical Aspects of Artificial Life

Mark A. Bedau*
Philosophy Department, Reed College
Portland, OR 97202-8199, USA
mab@reed.edu

October 7, 1996

Abstract

The field of artificial life (ALife) raises not only scientific but also philosophical questions, questions reaching into the heart of metaphysics, epistemology, and ethics. In part, ALife provides a fresh and rich context in which to consider longstanding philosophical concerns, but it also raises wholly new and distinctive questions. Some pertain to the practice of ALife science and a few concern intellectual bookkeeping, but the bulk are first-order questions about the nature of life and vital phenomena. The issues raised can be catalogued under five topics: (1) fundamental definitions, (2) artificial intelligence, (3) functionalism, (4) emergence, and (5) ethics. More than just raising these questions, reflection on ALife promises to provide the wherewithall to answer them, too. This paper clarifies fourteen questions and places them in the context of ALife research, with the aim of encouraging attempts to answer them.

There is a new interdisciplinary science—"artificial life" (or "ALife")—that promises to be of significant philosophical interest. Artificial life has relatively straightforward relevance to issues in metaphysics, philosophy of science, philosophy of biology, and philosophy of mind, but it also bears centrally on issues in social and political philosophy, economic philosophy, and ethics. However, few philosophers have begun to think through these issues, probably because few are yet even aware of this new science.

In order to help prompt reflection about the philosophical aspects of artificial life, this paper catalogues some of the philosophical issues ALife raises and

*For helpful comments, thanks to audiences at M.I.T. and the University of New Hampshire, and to Kate Elgin, Bernie Gert, Brian Goodwin, Mark Hinchliff, John Konkle, Nelson Minar, Jim Moor, Norman Packard, Steen Rasmussen, David Reeve, Walter Sinnott-Armstrong, and Carol Voeller.

potentially illuminates. The aim here is to describe and clarify the questions; answers are left for other occasions.

1 An Overview of Artificial Life

It is useful to contrast the field of artificial life with the analogous field of artificial intelligence (AI). Whereas AI attempts to devise and study computational models of cognitive processes such as reasoning, memory, and perception, ALife attempts to devise and study computationally implemented models of the processes characteristic of living systems. These processes include:

- self-organization, spontaneous generation of order, and cooperation
- self-reproduction and metabolization
- learning, adaptation, purposiveness, and evolution

Roughly speaking, what AI is to psychology and the philosophy of mind, ALife is to biology and the philosophy of biology.

But artificial life is more than just computationally oriented theoretical biology. It falls within a new branch of physics and computer science known as "complex systems theory," currently being presented in, e.g., the Addison-Wesley series SFI Studies in the Sciences of Complexity, and recently popularized through the science of chaos (Crutchfield et al. 1986, Gleick 1987). Living systems are especially salient examples of spontaneously organized complex systems, but nature exhibits many other examples as well—some outside the living realm, such as Bénard convection cells and certain other dissipative structures (Prigogine and Stengers 1984), and others consisting of aggregates of living beings, such as social groups or economic populations. The field of artificial life seeks an understanding of spontaneous organization and adaptation that illuminates *all* such phenomena.

In fact, one of the most striking natural systems that exhibits fundamental properties of life, such as self-organization and adaptive learning, is the mind. For this reason, the insights sought by artificial life may well shed light on at least some fundamental aspects of mental processes (a theme sounded in Bedau and Packard 1991 and in work on group "intelligence" in the social insects such as Collins 1991). ¿From this perspective, AI is a subfield within the larger ALife fold. And while life seems simpler than cognition, and ALife is concentrating its initial efforts on modeling simple living systems, many in the ALife community hope their work will eventually contribute to our understanding of mental processes.

The field of artificial life is betting on the success of a central working hypothesis: that *the essential nature of the fundamental principles of life can be captured in relatively simple models.* If correct, this hypothesis makes it sensible to attempt to understand the *general* class of living phenomena by computationally exploring the characteristic behavior of these simple models. Progress in ALife would then reveal fundamental truths about all living systems, including those found here on Earth. It is important to recognize that nothing guarantees that ALife's hypothesis is true. Thus, while the hypothesis does find some fledgling corroboration in work to-date, such as simple models of self-reproduction (Langton 1986) and adaptive evolution (Bedau and Packard 1991), its ultimate success is an empirical matter still hanging in the balance of future results.

ALife's working hypothesis provides a general motivation for the field's focus on computational models. Four more specific considerations bolster this rationale. First, ALife seeks maximally general principles governing all possible forms of life. So, rather than looking for regularities by sifting through a mass of data concerning the contingencies of life as we know it, it makes sense to study computational models since these exhibit an appropriately high level of generality. Second, it is very difficult, practically impossible, to conduct ordinary biological experiments that address the relevant questions even with respect to familiar forms of biological life. For example, most natural evolving systems operate on a time scale spanning vast numbers of human generations, and the few that evolve relatively quickly (like the immune system) are so complex that they all but defy analysis. By contrast, computer models allow for rapid and precisely controlled experiments. Third, the models explored by ALife exhibit quite complex forms of behavior, and in general it is simply impossible to discover their behavior by any means other than observing the effects of simulations (Wolfram 1984); in fact, this feature can be taken as definitive of complex systems. Fourth, abstract theorizing in the absence of empirical constraints is too easy, and the results are of uncertain value. A computer model must be precise if it is to be programmable, and all tacit assumptions must be explicitly specified if the program is to run. Implementing models on machines proves that the capacities captured by the model are mechanistically realizable. Furthermore, mechanistically implementing a model forces the scientist to surmount the non-trivial hurdle of computational feasibility (Garey and Johnson 1979). Thus, even if a computational model does not capture exactly how a certain kind of life process actually happens here on Earth, the model will be at least within the realm of feasibility.

Despite the analogies between artificial intelligence and artificial life, there is an important difference between the modeling strategies typically employed in the two fields. Most traditional AI models are top-down-specified serial systems involving a complicated, centralized controller which makes decisions based on access to all aspects of global state. The controller's decisions have the potential to affect directly any aspect of the whole system. However, most natural systems exhibiting complex autonomous behavior seem to be parallel, distributed networks of communicating "agents" making decisions that directly affect only their local state. Each agent's decisions are based on information about only its own local state. Following this lead, ALife is exploring the emergent dynamics of bottom-up-specified parallel systems of simple local agents.

Bottom-up models offer two important advantages over top-down models. First, the dramatic "toy" top-down models of early AI succumb to combinatorial explosion when attempts are made to have them produce natural, flexible, spontaneous behavior after being scaled up to realistic proportions (Dreyfus 1979, Hofstadter 1985). Bottom-up models, by contrast, have some promise of providing a computationally feasible solution to this kind of problem. Second, top-down models provide relatively shallow explanations of a system's global behavior. Rules governing the system's global behavior patterns are included in the system by fiat. Bottom-up models, by contrast, provide deeper explanations of a system's "macroscopic" behavior patterns by showing that they emerge out of the aggregate behavior of a population of "microscopic" agents.

ALife populations of processors have important similarities to parallel distributed processing networks (also known as PDP, connectionist, or neural networks) that today command so much philosophical attention (McClelland and Rumelhart 1986). Indeed, some ALife models include connectionist networks (e.g., Collins 1991). But most ALife models diverge from the connectionist paradigm in many specific respects, so familiar connectionist models constitute only a small subset of the class of bottom-up models being explored in artificial life.

Abstract accounts of artificial life like the foregoing have only limited value. A better taste of the real flavor of artificial life research can be had by reviewing typical ALife models such as those found in recent conference proceedings (Farmer et al. 1986, Langton 1989b, Langton et al. 1991).

2 Fourteen Questions

The field of artificial life raises a wide variety of philosophical questions. More importantly, an appreciation of artificial life might provide the wherewithal to find their answers. The following list, although not exhaustive, should convey the diversity of issues raised. Some are familiar philosophical questions that have interesting and rich instances concerning ALife; others are issues that arise only in the context of ALife. Some concern the nature of life and other vital phenomena; others concern the nature of artificial life science; a few concern intellectual bookkeeping.

The following five rough and overlapping topics serve to catalogue the questions: (1) fundamental definitions, (2) artificial intelligence, (3) functionalism, (4) emergence, and (5) the ethics of creation.

2.1 Fundamental Definitions

Philosophy is concerned with the essential aspects of the fundamental nature of reality; thus philosophy's preoccupation with such notions as existence, identity, causation, change, knowledge, mind, freedom, and value. One fundamental aspect of reality is *life*; the distinction between life and non-life is at least as basic as the distinction between mind and body. Yet philosophers have largely ignored the issue of the nature of life (Aristotle being one of the few exceptions). So, the most obvious philosophical issue raised by ALife work is perhaps:

Question 1: What is life?

Clarity about the nature of life should help clarify its possible subjects, so the answer to question 1 should help settle whether it could be appropriate to attribute life to systems consisting of ordinary biological individuals (such as species, ecosystems, and even the whole Earth), and to subsystems within biological organisms (such as the immune system). Closer to ALife's home, the answer to question 1 should also help settle the propriety of attributing life to a physical machine executing a suitable ALife program, to physical processes occurring inside that machine, or even to the abstract and disembodied program in its own right.

A skeptic might question whether there is any interesting answer to the question about life's fundamental nature, thinking the debate over the "true" nature of life to be mere semantics. Thankful to have progressed past nineteenth-century debates over vitalism, our skeptic might doubt that a principled distinction can be drawn in the field of more or less complex entities spanning crystals, viruses, bacteria, computer programs, and other self-organizing, adapting, evolving systems. But compare the issue of the nature of the mind. Few doubt that there is a real difference between mind and non-mind, yet the situations with respect to mind and life are strikingly similar. Just as there is a stong intuitive sense that mind is qualitatively different from non-mind, there is similarly a strong intuitive sense that life is qualitatively different from non-life. At the same time, mind and life both occur in various degrees; in fact, a case can be made for placing them at different ends of a hierarchy of internal processing strategies. These parallels provide ample reason for at least seeking a principled account for life.

Science has a tradition of attempts to provide a general characterization of life. Some proposals consist of more or less heterogeneous lists of characteristics (Monod 1972, Mayr 1982); others focus on specific properties, such as certain chemical patterns (Haldane 1949), evolvability (Maynard Smith 1975 and 1986), or autopoiesis, i.e., the property of being a "self-producing" system (Maturana and Varela 1980). Yet few biologists exhibit interest in the philosophical issue of the general nature of life. Their primary concern is to understand life-as-we-know-it, and an abstract philosophical account of life is largely irrelevant to that scientific pursuit. It is easy enough to identify life forms in our biosphere, so biologists quickly become engrossed in fruitful scientific projects such as describing and explaining patterns of diversity observed in biological populations and unravelling the structure and function of DNA.

ALife scientists have no comparable luxury. Striving to capture the general features characteristic of living systems, they must confront the question of the general nature of life-as-it-could-be. The salient features of life-as-we-know-it guide the formulation of initial hypotheses about the outlines of life-as-it-could-be, and life-as-we-know-it will remain an acid test of hypotheses about the most general nature of life. Nevertheless, we cannot sensibly design and evaluate ALife models without already having at least a rudimentary sense of the bounds of possible forms of life, and experimentation with ALife models will help refine our views on the general nature of life.

The targets of ALife models listed at the outset of section 1 (self-organization, metabolization, adaptation, etc.) are in effect an initial stab at partially characterizing the nature of life; thus these processes constitute at least a preliminary answer to the question about the nature of life (question 1). It is still a matter of debate whether all of these processes are equally fundamental to life and whether additional processes must be involved. But regardless of how these debates are resolved, ALife will remain interested in understanding these processes for their own sake. Thus, paralleling the question about the nature of life, ALife prompts us to ask:

Question 2: What is the nature of life's characteristic processes, such as self-organization, self-reproduction, metabolization, adaptation, purposiveness, and evolution?

These processes are undoubtedly related; understanding one will involve understanding the others. This raises a third question.

Question 3: How are life and each of its characteristic processes related, and what metaphysical status do these relationships have?

For example, is every extant form of life a member of an evolving system? If so, is this a mere contingency, a strict necessity, or a matter of some intermeditate modality?

It is also natural to wonder whether the fundamental definitions of concern to ALife are matters of a priori conceptual analysis, empirical data analysis, or something else. A central task of artificial life is to determine what can we learn about life and its characteristic processes from the structure of their ALife models, and this task raises a fundamental epistemological issue in the philosophy of science concerning scientific model building:

Question 4: What is the nature of the inference from the structure of a "successful" model to the properties of the phenomenon modeled?

It is a non-trivial matter to determine which features of a model are necessary or sufficient aspects of reality. For example, if bottom-up models persist as the architecture of choice in both AI and ALife, it is unclear whether to attribute this to the nature of life and mind or to our own epistemological limitations.

However we are to account for the epistemological significance of ALife models, there is clearly a fertile dialectical relationship between our grasp of the answer to this philosophical issue and the models produced and studied in artificial life. ALife models that attempt to capture vital processes are motivated and directed by an antecedent intuition about the nature of these processes, coupled with an antecedent conviction about their centrality to life. However, the successful construction of models might make it apparent that the true nature of these processes or the role they play in living systems is *not* as previously presumed. Model building can highlight the need for revision in our conceptions of the nature of these processes and of life. Furthermore, appreciation of the model can suggest the direction and substance of these revisions. These revised views can generate a new cycle of ALife model building, which can prompt yet further revisions in our understanding of life processes. Thus cycles the dialectic between the philosophy and science of ALife.

2.2 Artificial Intelligence

Artificial intelligence has profoundly affected contemporary philosophy. Much attention has been given to weighing its philosophical implications (Boden 1977 and 1990, Dennett 1978, Dreyfus 1979, Haugeland 1985, Hofstadter 1985). Although many conclusions are still controversial, there is wide agreement about which issues are important and which positions deserve serious attention.

Since ALife and AI share so many important features, we can exploit the familiarity of the intellectual terrain around AI to identify some of the main landmarks around ALife, and initial attempts to do this have already begun (Sober 1991). At the same time, we must be prepared for the possibility that more extensive exploration of ALife might force us to revise what we now take to be settled conclusions about AI.

We noted above that traditional AI has certainly not enjoyed overwhelming and unequivocal success, because of combinatorial explosion and the "frame" problem (Dreyfus 1979, Haugeland 1985, Hofstadter 1985, Boden 1990). These very problems show some promise of being mitigated by the bottom-up modeling strategies embodied in the new PDP models (McClelland and Rumelhart 1986). Although it is too early to judge the final success of the PDP approach, the similarity between PDP models and the bottom-up models found in ALife is still quite striking. This raises the bookkeeping issue of how to classify ALife models:

Question 5: What is the relationship between typical ALife models and connectionist (PDP, neural network) models?

It is important to note that ALife and connectionism pertain to different *kinds* of classifications. Whether a model is connectionist concerns its architecture, and connectionist architectures can be applied to problems concerning widely different kinds of phenomena. What makes a model ALife, on the other hand, is the kind of phenomenon being modeled, and widely different kinds of architecture can be employed in ALife models.

Nevertheless, it turns out that ALife models almost invariably possess a certain broad architectural feature—being "bottom-up"—which they share with connectionist models. For our purposes, bottom-up models are those in which the global behavior statistically emerges from the aggregate behavior of a population of relatively simple and autonomous processors acting solely on the basis of local state information. Connectionist models are a specific subset of bottom-up models in which the population of processors is a set of nodes arranged in layers, where a node's activity is some simple function of its input from the nodes to which it has immediate connections, and the strengths of the connections between nodes are adjusted in learning. Most ALife bottom-up models lack such specifically connectionist architectural details. Part of what makes this bookkeeping question interesting is the possibility of learning something about the significance and limitations of connectionism by reflecting on the architecture and performance of ALife models. But its main interest, surely, comes from the prospect of learning something fundamental about life from the invariant underlying architecture of its models.

Another kind of issue is raised by the close relationship between artificial life and artificial intelligence. They not only use the same methodology; many of life's fundamental principles seem true also of mental processes. This blurs the disciplinary relationship between AI and ALife, and raises a basic question:

Question 6: What is the relationship between life and mind?

Are life and mind two endpoints on a continuum? Different stages in some hierarchy? What fundamental principles apply to both, and why? One of the most exciting prospects of ALife is that experimental research can be brought to bear on the connection between these two most basic aspects of reality.

Of the many philosophical controversies involving artificial intelligence with analogues concerning artificial life, one deserves special mention. Usually a computer model of something can easily be distinguished from the real thing. A hurricane in a computer simulation of the weather is not a real hurricane, and even though modern flight simulators are strikingly realistic a session in a flight simulator is not really flying. But in some cases a suitable computer simulation *does* seem to produce the genuine article (Dennett 1981). A suitable computer simulation of jazz improvisation seems for all the world to produce real jazz music, and a computer-driven theorem-prover seems to produce proofs every bit as genuine as those produced by flesh-and-blood mathematicians.

One of the most heated philosophical controversies raised by AI is the question whether it is ever appropriate to attribute mental properties such as thinking to computers (or programs). The starkest form of this issue is this: Even if we assume that it is possible for a computer (perhaps in the future) to *simulate perfectly* the behavioral and functional processes characteristic of mental beings, does it follow that the computer *really* has a mind? Alan Turing (1951) initiated this debate by proposing the Turing test in his seminal argument for a positive answer, and the debate has recently been revived by John Searle's (1980) "Chinese room" argument for a negative answer.

One of the more dramatic philosophical issues raised by ALife is the parallel question about whether a perfect simulation of life would be the creation of real life:

Question 7: Under what conditions, if any, is a simulation of a life process an artificial but real instance of life?

What makes ALife so exciting to the mass media is this prospect of creating new forms of life. It is certainly possible that distant galaxies contain forms of life quite unlike those with which we are now familiar. Why couldn't quite different forms of life be created in artificial life laboratories? The simulation-or-reality debate in artificial life has already been engaged, with Christopher Langton (1989a) arguing the positive position and H. H. Pattee (1989) raising skepticism.

ALife's simulation-or-reality debate bears on other philosophical issues. For example, Langton's argument would be significantly supported by a functionalist approach to life (discussed in the next section). Furthermore, the simulation-or-reality debate seems more tractable in artificial life than in artificial intelligence because ALife can sidestep some of AI's sharpest thorns—life need not involve subjectivity and self-consciousness, for example (so there is no evident analogue of Searle's Chinese room argument for artificial life), and the prospect of artificially created life apparently threatens our self-esteem much less than the prospect of artificially created minds. Progress on ALife's simulation-or-reality debate might even help break the impasse in the analogous debate in AI.

2.3 Functionalism

Functionalism is the dominant position in contemporary philosophy of mind (Putnam 1975, Fodor 1981). Functionalists view mental beings as input-output devices and hold that having a mind is no more and no less than having a set of internal states that causally interact (or "function") with each other and with environmental inputs and behavioral outputs in a characteristic way; a mental system is any system whatsoever that is governed by a set of internal states that is functionally isomorphic to human mental states. It does not matter what kind of material instantiates those functionally-defined patterns. Human mental states happen to be embodied in patterns of neuronal activity, but if exactly the same patterns of activity were found in a system composed of quite different materials—such as silicon circuitry—then, according to functionalism, that system would also literally have a mind. Functionalism's slogan could be "mind as software, not hardware," and its central thesis could be summarized thus: Mind is a property of *form*, not matter; to have a mind is to embody a distinctive dynamic *pattern* of states, not to be composed out of a distinctive sort of substance.

Life, too, seems to hinge on form rather than matter. Living organisms participate in a network of processes, some (such as information processing, metabolization, purposeful activity) operating within the organism's lifetime, and others (such as self-reproduction and adaptive evolution) operating over many generations. These processes must occur in some material substratum or other, but which specific kind of matter embodies them seems irrelevant to a system's vitality so long as the *forms* of the processes are preserved. The universe may well contain alien forms of life in which vital processes are sustained in substances outside even the realm of carbon chemistry. Thus, just as with the mind, functionalism is an attractive approach toward the nature of life, which raises the question:

Question 8: Is life a functional notion?

A functionalist perspective on life, if sound, is important aside from the intrinsic interest of understanding life. The contrast afforded by a functional definition of life would potentially illuminate functionalism with respect to mind. Furthermore, the notion of functional definition in general is interesting, but so far all the best examples of functional definitions have come from the philosophy of mind. Providing a second rich and complex setting in which to explore functional definitions could not help but shed light on functionalism in general.

Just as functionalism in the philosophy of mind has been deeply informed by its connection with artificial intelligence, functionalism with respect to life should have a similarly close association with artificial life. In the previous section we noted that ALife models may well transform artificial intelligence, partly by supporting and extending the connectionist revolution and partly by connecting the studies of mind and life. If AI does undergo such transformations, then the face of its intellectual cohort—functionalism in the philosophy of mind—might be similarly transformed. This highlights the need for a grasp of the different kinds of accounts that fall under the functionalist fold:

Question 9: What are the different basic kinds of functional accounts?

We need to understand how the form of a functional account of some phenomenon depends on the architecture of our best models of that phenomenon. The bottom-up architecture of ALife models might especially transform functionalism. For example, it has been argued that functionalism must explicitly acknowledge that the global "macroscopic" patterns definitive of certain functional systems emerge bottom-up style from an underlying "microscopic" layer of phenomena (Bedau 1991). Awareness of ALife models forces us to reflect on the relationship between lower-level and higher-level features of emergent functionalist systems and rethink which aspects of a type of phenomena functionalist models must capture. In this way we are led to confront a fundamental epistemological issue concerning functionalism:

Question 10: What are the criteria of adequacy for emergent functionalist accounts?

Although functionalism has attracted attention primarily within the philosophy of mind, its appeal extends to social and political philosophy, economic philosophy, and the special sciences generally (Fodor 1981). To the extent that ALife does change the face of functionalism, the ramifications will extend potentially to all the special sciences. The lessons about functionalism learned from ALife models might help us understand such diverse subjects as the complex patterns observed in voting behaviors and economic transactions.

2.4 Emergence

The issue of emergence has been epitomized by the question: "How can the whole be greater than the sum of its parts?" This issue arises because in certain situations complicated phenomena do seem to emerge spontaneously from fundamentally simpler phenomena.

Some examples of apparent emergence involve non-living matter. A hurricane is an autonomous, self-sustaining global unity with an integrated dynamics, but it somehow emerges simply from the aggregate action of a huge collection of air and water molecules all governed by the same relatively simple local rules. Another example of apparent emergence is the mind. Conscious mental life consists of an autonomous, integrated flow of mental states (beliefs, desires, etc.) that follows a complicated global dynamic, but presumably this process somehow emerges out of the aggregate activity of the huge interconnected collection of neurons in the brain that are all obeying more or less the same relatively simple local rules.

Life is the context for many of the most compelling examples of apparent emergence. *The origin of life.* The initial primitive biotic community of self-replicating adaptively evolving entities emerged from a prebiotic chemical soup. *Phylogeny.* A vast biosphere containing a multitude of different complex species has emerged in the course of evolution from an initial biosphere containing just a few relatively simple species. *Ontogeny.* An individual organism is a vast diversity of systems and subsystems composed of a vast diversity of specialized types of cells. This intricately differentiated entity emerges in the course of embryological development from a single undifferentiated zygote. *The vital hierarchy.* Ecosystems are composed of organisms, which are composed of organ systems, which are composed of organs, which are composed of tissues, which are composed of cells, which are composed of organelles, which are composed of chemicals. Each "layer" in this hierarchy consists of two "levels"—a macro-level the behavior of which somehow emerges from the aggregate behavior at the micro-level below it.

What to make of emergent phenomena has been a perennial philosophical problem, one so far resisting scrutiny. Vitalistic explanations employing occult substances or forces operating outside the realm of physical and chemical law are unacceptable. Traditional forms of reductionism are unappealing explanations of emergence since the necessary reductions seem unavailable. Even though the micro-level entities behave mechanistically and the system's global macro-level behavior that emerges is driven solely by the micro-level dynamics, without explicit micro-level simulation the global behavior cannot be predicted even from complete knowledge of the micro-level properties—in this sense the emergent behavior is autonomous. Contemporary philosophy, deeply suspicious about whether the idea of emergence is even coherent, has simply put this issue on indefinite hold.

Artificial life is squarely attempting to understand the emergent quality of vital phenomena, and bottom-up ALife models typically display forms of emergent behavior. Brief descriptions of two simple models of group behavior might begin to suggest the emergent potential of bottom-up models. Although not themselves models of systems that are alive, these models do exhibit the emergence of simple forms of some of the central characteristics of living systems, such as spontaneous self-adaptivity and self-organization.

Flocking Boids. Flocks of birds exhibit impressive group behavior, fluidly swooping, reeling and maneuvering around obstacles, all the while maintaining group cohesion while avoiding collisions. Furthermore, the flock's dynamic equilibrium is maintained without any individual bird or group of birds functioning as a leader. A recent computer model of flocking "boids" developed by Craig Reynolds (1987) has reproduced similar behavior in simulated flocks. What is interesting is that the natural flocking behavior emerges spontaneously from models that are driven simply by each individual boid independently determining its flight path from local information, following such rules as match the velocities of neighboring boids, steer towards the perceived center of mass of neighboring boids, and minimize (within limits) the distance from neighboring boids. Simulations show that when a random assortment of individuals boids follow these local rules, natural flocking behavior arises—the flock coheres and moves as a unit, fluidly changing direction and navigating arbitrarily placed obstacles. (A sequence of pictures showing flocking boids is reproduced in Langton 1989a.) Even though the global behavior of this model is driven solely by the simple local rules governing each individual boid, global flocking behavior spontaneously emerges.

The Game of Life. Probably the best known example of something like an A-Life model is the Game of Life devised more than a generation ago by the Cambridge mathematician John Conway (1982) and popularized by Martin Gardner (1983). "Played" on a two-dimensional grid (such as a checker board), at each time step each square or cell is in one of two states—"dead" or "alive."

Whether a given cell is dead or alive at a given time is a simple function of the previous states of the eight adjacent cells: a cell that was alive at t remains alive at $t+1$ if and only if exactly two or three of its neighbors were alive at t (cells with fewer than two living neighbors die of "loneliness" and those with more than three die of "overcrowding"); a cell that was dead at t becomes alive at $t+1$ if and only if exactly three of its neighbors were alive at t (there were just enough living neighbors to "breed" a new living cell). Extensive simulations have shown that amazingly complex patterns can emerge from this simple local rule governing individual sites: some reach stable or periodically oscillating configurations while others continue to change and grow indefinitely. Clusters of cells can function just like AND, OR, NOT, and other logic gates, and these gates can be connected into complicated switching circuits. They can even constitute a universal Turing machine.

Reynold's boids, Conway's Game of Life, and more complicated artificial life models make it possible to reopen the study of emergence. They provide a profitable setting for gathering rich and manageable empirical data from models of different kinds of emergent phenomena, thus providing a new purchase on the question:

Question 11: Under what conditions do systems exhibit emergent properties?

This question applies to both artificial and natural systems, and ALife might help answer both. First, as outlined above, many aspects of life apparently involve emergent phenomena, so any fundamental understanding of living systems provided by ALife has the prospect of illuminating the general properties of emergence. But more importantly, bottom-up ALife models themselves illustrate and instantiate emergent dynamics. It is primarily for the latter reason that ALife might provide a key which unlocks the mysteries of emergence. The hope is that understanding how ALife models generate emergent phenomena will reveal how emergent phenomena arise in real living systems.

One preliminary task for the study of emergence in ALife models is to formulate a typology of basic kinds of emergent phenomena. This and other fruits of the study of ALife emergence should help sort out a philosophical bookkeeping question:

Question 12: How is emergence related to reduction, supervenience, explanation, prediction, and determinism?

Progress on this sort of question will ultimately require a precise account of bottom-up models—something that it must be admitted is still lacking. Study of ALife emergence can be expected to help settle whether the difference between bottom-up and top-down models is a matter of principle, degree, perspective, or something else.

2.5 The Ethics of Creation

Artificial life raises two kinds of ethical issues. One is a concern about the consequences of technological change for extant beings; the other concerns the consequences of technological change for newly created life forms.

The development of a powerful new technology usually has ethical consequences. Atomic fission and fusion provide examples from recent history, and today genetic engineering threatens to teach us of this lesson again. One kind of ethical issue raised by ALife is analogous to those raised by other powerful technologies—details differ but the form is similar. The havoc wreaked by computer "viruses" and "worms" (Spafford 1991) provide ample evidence of the destructive potential of artificial life technologies; precautions must be taken against their accidental misuse and intentional abuse. These concerns highlight a question of the ethics of technology:

Question 13: What ethical implications does ALife research have for humanity and other extant forms of life?

Reflection on this topic will identify ethical predicaments which ALife scientists must confront. (It is ironic that ALife might provide *solutions* to some of the very problems it spawns; e.g., there is talk about creating artificial immune systems that protect computer networks from new viruses.)

Technologies for creating life involve a special ethically-charged consequence: the creation of new living beings. Consider genetic engineering. If genetic engineering produces complex enough forms of life, ethical consequences *for the newly created forms of life* might arise. To express this issue at its logical if fanciful extreme, we would be wrong to pretend we had no ethical responsibility to a future Frankenstein we created by genetic engineering. Now, this digression on genetic engineering is beside the point if the science of artificial life concerns only certain computer models and their theoretical implications. However, *if* the simulation-or-reality issue about life (question 7) has a positive answer, then artificial life *computer* laboratories might witness the creation of new forms of life just as might the bio-chemistry laboratories of genetic engineering. The "creatures" being simulated in a suitable ALife model (sufficiently elaborate, sufficiently long-operating, ...?) could actually be alive. Thus, pending the outcome of question 7, ALife raises a more pointed ethical issue:

Question 14: What ethical responsibilities would we bear to artificial forms of life that we created?

Presumably the impermissibility of whimsically harming or destroying creatures would not disappear merely if the "creatures" were created by genetic engineering or artificial life techniques. Most of us do not think twice about swatting a mosquito whereas we hesitate at even kicking a dog. It is sometimes thought that our moral responsibilities to a creature depend on which specific capacities it possesses. The animal rights literature, for example, bases our responsibilities to animals on their sentience—in particular, their capacity to experience pleasure and pain (Regan 1983). The environmental ethics literature, by contrast, tends to ground our responsibilities to other forms of life on the supposition that life in-and-of-itself is intrinsically valuable (Callitott 1982, Taylor 1986). Artificial life might create a new setting in which fundamental ethical issues like these must be pondered.

3 Conclusion

This list of questions should make amply evident the philosophical interest of the new science of artificial life. ALife provides a new and distinctive context in which to consider longstanding philosophical concerns, and it raises wholly new and distinctive questions. These questions reach into the heart of metaphysics, epistemology, and ethics.

Furthermore, reflection on ALife science might suggest how to answer these questions. The scientific strategies being developed in ALife are opening new philosophical horizons. In time, artificial life may well affect the substance of philosophy as much as, if not more than, artificial intelligence.

The philosophy of artificial life is not merely a derivative, second-order gloss on first-order ALife science. In fact, most of the questions identified above are of direct and fundamental concern in artificial life *science*. However unclear the relationship between ALife science and ALife philosophy might be, the two are surely closely related, and in time they will co-evolve.

References

Bedau, M. A. 1991. “Emergent Functionalism, Artificial Intelligence, and Artificial Life.” In preparation.

Bedau, M. A., and N. H. Packard. 1991. “Measurement of Evolutionary Activity, Teleology, and Life.” In *Artificial Life II*, edited by C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen. SFI Studies in the Sciences of Complexity, Vol. X. Redwood City, CA: Addison-Wesley.

Berlekamp, E. R., J. H. Conway, and R. K. Guy. 1982. *Winning Ways*. Vol. 2. New York: Academic Press.

Boden, M. A. 1977. *Artificial Intelligence and Natural Man*. New York: Basic Books.

Boden, M. A., ed. 1990. *The Philosophy of Artificial Intelligence*. New York: Oxford University Press.

Callicott, J. B. 1989. *In Defense of the Land Ethic.* Albany: SUNY Press.

Collins, R. J., and D. R. Jefferson. 1991. "AntFarm: Towards Simulated Evolution." In *Artificial Life II*, edited by C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen. SFI Studies in the Sciences of Complexity, Vol. X. Redwood City, CA: Addison-Wesley.

Crutchfield, J. P., J. D. Farmer, N. H. Packard, and R. S. Shaw. "Chaos." *Scientific American*, December 1986, 46–57.

Dennett, D. C. 1978. "Artificial Intelligence as Philosophy and as Psychology." In his *Brainstorms.* Montgomery, VT: Bradford Books.

Dennett, D. C. 1981. "Reflections [on D. R. Hofstadter's 'The Turing Test: A Coffeehouse Conversation']." In *The Mind's I*, edited by D. R. Hofstadter and D. C. Dennett. New York: Bantam Books.

Dreyfus, H. L. 1979. *What Computers Can't Do.* 2nd edition. New York: Harper and Row.

Farmer, D., A. Lapedes, N. Packard, and B. Wendroff. 1986. *Evolution, Games, and Learning: Models for Adaptation in Machines and Nature.* Amsterdam: North-Holland.

Fodor, J. A. 1981. *Representations.* Cambridge, MA: The MIT Press.

Gardner, M. 1983. *Wheels, Life, and Other Mathematical Amusements.* New York: Freeman.

Garey, M. R., and D. S. Johnson. *Computers and Intractability.* New York: Freeman.

Gleick, J. 1987. *Chaos: Making a New Science.* New York: Viking.

Haldane, J. B. S. 1949. *What is Life?* London: Alcuin Press.

Haugeland, J. 1985. *Artificial Intelligence: The Very Idea.* Cambridge, MA: The MIT Press.

Hofstadter, D. R. 1985. "Waking Up from the Boolean Dream, *or*, Subcognition as Computation." In his *Metamagical Themas: Questing for the Essence of Mind and Pattern.* New York: Basic Books.

Langton, C. G. 1986. "Studying Artificial Life with Cellular Automata." In *Evolution, Games, and Learning: Models for Adaptation in Machines and Nature*, edited by D. Farmer, A. Lapedes, N. Packard, and B. Wendroff. Amsterdam: North-Holland.

Langton, C. G. 1989a. "Artificial Life." In *Artificial Life*, edited by C. Langton. SFI Studies in the Sciences of Complexity, Vol. VI. Redwood City, CA: Addison-Wesley.

Langton, C. G., ed. 1989b. *Artificial Life*. SFI Studies in the Sciences of Complexity, Vol. VI. Redwood City, CA: Addison-Wesley.

Langton, C. G., C. E. Taylor, J. D. Farmer, and S. Rasmussen, eds. 1991. *Artificial Life II*. SFI Studies in the Sciences of Complexity, Vol. X. Redwood City, CA: Addison-Wesley.

Maturana, H. R., and F. J. Varela. 1980. *Autopoiesis and Cognition*. Dordrecht: Reidel.

Maynard Smith, J. 1975. *The Theory of Evolution*. 3rd edition. Hammondsworth: Penguin.

Maynard Smith, J. 1986. *The Problems of Biology*. New York: Oxford University Press.

Mayr, E. 1982. *The Growth of Biological Thought*. Cambridge, MA: Harvard University Press.

McClelland, J. L., and D. E. Rumelhart. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. 2 Vols. Cambridge, MA: The MIT Press.

Monod, J. 1972. *Chance and Necessity*. Translated by A. Wainhouse. New York: Vintage.

Pattee, H. H. 1989. "Simulation, Realizations, and Theories of Life." In *Artificial Life*, edited by C. Langton. SFI Studies in the Sciences of Complexity, Vol. VI. Redwood City, CA: Addison-Wesley.

Prigogine, I., and I. Stengers. 1984. *Order out of Chaos*. Toronto: Bantam Books.

Putnam, H. 1975. "The Nature of Mental States." In his *Mind, Language, and*

Reality. Cambridge: Cambridge University Press.

Regan, T. 1983. *The Case for Animal Rights*. Berkeley: University of California Press.

Reynolds, C. W. 1987. "Flocks, Herds, and Schools: A Distributed Behavioral Model." *Computer Graphics* 21: 25–34.

Searle, J. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3: 417–458.

Sober, E. 1991. "Learning from Functionalism." In *Artificial Life II*, edited by C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen. SFI Studies in the Sciences of Complexity, Vol. X. Redwood City, CA: Addison-Wesley.

Spafford, E. H. "The Internet Worm: Crisis and Aftermath." *Communications of the ACM* 32: 678–687.

Taylor, P. 1986. *Respect for Nature*. Princeton: Princeton University Press.

Turing, A. M. 1951. "Can a Machine Think?" *Mind* 59: 433–460.

Wolfram, S. "Computer Software in Science and Mathematics." *Scientific American*, September 1984, 188–203.