DOI: http://dx.doi.org/10.7551/978-0-262-32621-6-ch096

Artificial Life and the Philosophy of Science

Mostafa Fahmy

Carnegie Mellon University, Pittsburgh, PA 15213
mfahmy@cmu.edu

Abstract

What is the ontological status of Artificial Life systems? This question admits two interpretations. The first asks whether and to what degree Artificial Life systems could potentially be considered 'alive' in the sense that we take naturally occuring biological systems to be. The second has to do with what sort of role Artificial Life systems occupy in scientific inquiry. In this paper, we investigate both facets of the question, and explore the philosophical and practical implications of possible answers for scientific practice.

Introduction

The contemporary conception of Artificial Life as a discipline was put forward by Christopher Langton. Langton (1993) characterized the field broadly as "devoted to studying the scientific, technological, artistic, philosophical, and social implications" of creating "living" artifacts.

For many of its practitioners, the ultimate goal of Artificial Life is the synthesis of systems that exhibit life-like behavior, organization, and complexity. In engineering and evolving such systems, we hope to gain insight into the dynamics which govern the natural phenomena that inspire them.

Artificial Life raises a number of compelling issues for the philosophy of science. One of the most pertinent concerns can be stated as follows: what is the ontological status of Artificial Life systems?

This question admits two interpretations. The first, a broad interpretation which we shall call *the metaphysical question*, has to do with the place of Artificial Life systems in our ontology of the world at large. Specifically, it asks whether and to what degree Artificial Life systems could potentially be considered 'alive' in the sense that we take naturally occuring biological systems to be. The second, a narrow interpretation which we shall call *the methodological question*, has to do with what sort of role Artificial Life systems occupy in the context of scientific inquiry. Are they theoretical entities, objects of experiment, or something altogether different?

The present paper seeks to explore both interpretations. First, we will trace the development of a particular strain of ideas from the Western philosophical canon to inform our understanding of the metaphysical question, as well as the roots of Artificial Life as a discipline. Some scientific and ethical implications of potential answers to the metaphysical question will be discussed.

Second, we will survey a variety of perspectives on the role that Artificial Life systems play, or could potentially play, in scientific practice. This portion of our exploration will focus primarily on software-based Artificial Life.

Philosophical Motivations for Artificial Life

Proponents of the Artificial Life (hereafter ALife) program claim that the construction of living or life-like artifacts can help in addressing fundamental problems and open questions in the science and philosophy of biology. It will serve our inquiry to begin by considering some possible motivations for this claim. Three such motivations are discussed below.

The Generalization Problem

Biologists and philosophers of biology are interested in deriving general principles and theories regarding living systems. Langton (1989) argues that this task requires identifying the essential properties of living systems and separating them from those that are accidental.

A fundamental obstacle to this endeavor is that all naturally occurring forms of life share a common origin. This makes it difficult, if not impossible, to differentiate between those features which are shared among living systems because they are essential, and those that are universal (due to common origin) but accidental. This presents a problem for anyone wishing to form general theories about life which capture only its essential properties.

One way of overcoming this predicament is to search the cosmos for extraterrestrial forms of life (Smith, 1992). Another option is to attempt to construct new forms of life ourselves. In doing so, we would effectively be exploring "the biology of possible life" (Langton, 1989):

> Only when we are able to view *life-as-we-know-it* in the larger context of *life-as-it-could-be* will we really

understand the nature of the beast.

Langton's argument is based on an Aristotelean conception of science, one which entails a metaphysical committed to the existence of one or more *natural kinds* - natural (i.e. human-independent) groupings of particular objects - elements of which are said to share a privileged set of properties (their *essence*) without which they would cease to be objects of that kind. Specifically, Langton assumes that (1) 'life' is a natural kind, and (2) an entity is said to be alive by virtue of possessing a specific set of essential properties shared by all living things.

The commitment to natural kinds is common among scientific realists - those who take science to be representative of, or at least aiming to uncover, the actual workings of the world. It is Langton's second commitment, the commitment to essences, which casts him as an Aristotelean.

For the non-Aristotelean, there are at least two other possible motivations for the ALife program.

Understanding Through Synthesis

Richard Feynman is said to have written, "What I cannot create, I do not understand." (CaltechArchives, 1988).

Scientists and philosophers of biology are interested not only in describing the behavior of living systems, but also in explaining how that behavior is generated. Attempting to synthesize life-like systems provides us with an opportunity to test our understanding of the dynamics that account for the behavior we observe in living systems.

Overcoming Practical Limitations

Occasionally, there exist real-world obstacles which preclude scientists from carrying out certain types of inquiries. These obstacles may be related to limited budgets, technological constraints, or ethical reservations.

In such situations, computer modeling and simulation can offer an attractive alternative to traditional experimental methods. In the context of biology, this often results in the construction of systems which fall under the purview of Artificial Life.

The Metaphysical Question

There are two primary theses regarding the metaphysical status of artificial life (hereafter, ALife) systems:

The *Strong ALife* thesis is the claim that life is a process which can be abstracted away from any particular medium.

The *Weak ALife* thesis denies the possibility of creating living artifacts, but claims that the mechanisms and dynamics which govern living systems can be studied by constructing artifacts which model or simulate such systems.

There is also a third, more pragmatically oriented thesis, which is orthogonal to the Strong/Weak ALife debate. *Instrumental ALife* simply states that ALife systems can serve as tools for solving practical problems.

The Strong ALife and Weak ALife positions parallel the Strong AI and Weak AI theses, respectively. In artificial intelligence and computational psychology, Strong AI is the claim that it is possible to program computers so that they exhibit general intelligence of the same or similar kind that humans possess. In other words, "the appropriately programmed computer is [a] mind and has intentions" (Searle, 1980). Conversely, Weak AI denies the possibility that computers can think, but claims that they can be programmed to act as though they were intelligent in narrow scopes.

Proponents of Strong AI often ground their position in a functionalist theory of mind. On this view, it is not the underlying structural economy of parts which accounts for intelligence. Instead, it is the functional organization of those parts that is key. Thus, mental states are multiply realizable —the same mental states that occur in the brain can be instantiated in other media so long as the parts of the system are organized in such a way as to be functionally equivalent.

Similarly, Strong ALife can be seen as advocating a functionalist theory of life. The claim is that non-biochemical systems can be organized in such a way that their behavior is functionally equivalent (or sufficiently similar) to that of naturally occurring living systems. On this view, since the appropriate level of description for the properties of life is functional (and thus divorcable from considerations of medium), it follows that any system which satisfies these functional properties is in fact genuinely alive. Put another way, life itself is multiply realizable.

How did we come to arrive at such a view of life? In the next section of this paper, we will trace the history of philosophical ideas which serve as precursors to the contemporary conception of Artificial Life.

The Philosophical Roots of Artificial Life

Ancient Materialism The *physiologoi* (literally *physical* or *natural philosophers*) were the prototypical scientists of Ancient Greece. They perceived lightning in the sky and saw not the manifestation of Zeus' anger, but a natural process to be studied and understood. Among them was Democritus, one of the early atomists. The atomists held that the universe was composed of indivisible atoms whose movements and interactions took place in an infinite void.

Democritus believed that all living things possessed a *psyche*, or soul, which was the feature that allowed them to carry out their life functions. This psyche was thought to be composed exclusively of fire atoms (Berryman, 2010). As archaic as this idea seems today, it is significant in that Democritus' was one of the first materialist accounts of life. It is reductionist, in that it claimed that the property of life could be reduced to and explained entirely in terms of interactions of matter.

Modern Mechanism In the 17th Century, the scientific revolution was sweeping through Europe. One prominent strain of thought that emerged during this period was the idea that the natural world and its parts could be seen as complex machines, similar to intricate watches and automata that were popular at the time. This idea, that we live in a sort of clockwork universe, was called *mechanism*, or the *mechanical philosophy*.

The French philosopher and mathematician René Descartes partially revived Democritus' idea that life had a material basis. Descartes claimed that animals were nothing more than non-sentient automata (Harrison, 1992). Humans were distinct from other animals in that they had minds, which were constituted of a different substance than matter (a position known as Cartesian dualism). Other animals, however, lacked minds —they were merely physical machines. Thus their behavior could be entirely accounted for by the interaction between their internal parts and the influence of outside forces on those parts. This applies to human bodies as well, insofar as they are physical (though the mind, being nonmaterial, merits different considerations). In his 1664 *Treatise on Man*, Descartes (1985) describes a hypothetical construct, an artificial man whose internal constitution and outward behavior is meant to mimic our own. Descartes concludes:

> I should like you to consider, after this, all the functions I have ascribed to this machine [...] I should like you to consider that these functions follow from the mere arrangement of the machine's organs every bit as naturally as the movements of a clock or other automaton follow from the arrangement of its counter-weights and wheels.

The English philosopher Thomas Hobbes, one of Descartes' contemporaries, was an even stauncher materialist. Hobbes is primarily known for his contributions to political philosophy and his computational approach to psychology. Hobbes held that the human faculty of reason was nothing but 'reckoning' —that is, calculation or computation. This being the case, the behavior of man could be explained and predicted if the elements of this computation (the dispositions and forces driving man's actions, and their combinations) could be identified. Hobbes employs this mechanistic understanding of human psychology to provide a proto-game-theoretic account of man's transition from the state of nature (man's pre-societal condition) to the formation of a civil state. This account is given in his seminal work, *Leviathan*. Hobbes (1651) opens with the following observation:

> NATURE (the art whereby God hath made and governs the world) is by the art of man, as in many other things, so in this also imitated, that it can make an artificial animal. For seeing life is but a motion of limbs, the beginning whereof is in some principal part within, why may we not say that all automata (engines that move themselves by springs and wheels as doth a watch) have an Artificial Life?

It is interesting to note that Hobbes' introduction mirrors that of Langton, in an updated version of his original *Artificial Life* paper (Langton, 1993):

> "Art" + "Life" = Artificial Life: Life made by Man rather than by Nature. Our technological capabilities have brought us to the point where we are on the verge of creating "living" artifacts.

The Darwinian Revolution In *On the Origin of Species*, Charles Darwin (1859) described the process of evolution by natural selection. Darwin argued that the myriad forms of life we observe are not static, but the result of a process of branching evolution taking place over many generations.

Scientists and philosophers are interested in providing explanations which address phenomena at an appropriate level of description. A level of description that is too low results in an explanation that is difficult to work with. One that is too high obscures the insights one might gain from a more appropriate level of description. One of Darwin's revolutionary insights, which Dennett (1995) calls Darwin's Dangerous Idea, is that the algorithmic level is the level that best accounts for the variation and qualities of life we observe.

The Functionalist Turn In the late 1940's, John von Neumann was investigating the possibility of constructing a self-replicating machine. He undertook the construction of such an automata, described kinematically (with reference to aspects of motion apart from considerations of mass and force), as a thought experiment. Levy (1992) recounts the story of the attempt as follows.

Von Neumann posited a creature living on a lake containing the same elements of which the creature was composed. He went on to describe the information- and material-processing parts of the creature and how these could be combined in such a way as to provide structures by which the creature could, by manipulating the available elements in its environment, construct a copy of itself.

This attempt was widely met with skepticism. For though the process von Neumann described was taken to be sound, his peers thought the origin of such an automata to be too opaque. Von Neumann had posited, in their minds, a black box. One of von Neumann's colleagues, a mathematician named Stanislaw Ulam, suggested to von Neumann do away with the material environment which his creature inhabited and instead place it on an abstract grid, wherein its workings could be described functionally. This led to the birth of the cellular automaton and the publication of *The General and Logical Theory of Automata*, in which Von Neumann (1951) stressed the analogies between living organisms and his digital automata.

The Birth of ALife With Democritus, we see one of the first materialist accounts of life. Centuries later, his ideas are picked up and extended by mechanists like Hobbes and Descartes. They argue that living things are physical automata whose overall behavior is explained by the interactions of their constituent parts and external physical forces. Darwin provides us with an account of life in which the algorithmic processes of evolution by natural selection plays a fundamental role in shaping the variety of mechanisms we observe in living things. In the 20th century, von Neumann abstracts the process of life away from considerations of underlying medium. The character of the narrative is one that begins with a material reductionism of life, which is then given a mechanistic treatment, and ultimately abstracted to yield a functionalist account of life.

This pattern of ideas culminates in the organization of the first Artificial Life conference by Christopher Langton, who also publishes the manifesto that lays out the vision of the ALife program (Langton, 1989). Langton's conception of life, and of Strong ALife, is one that is both "unashamedly functionalist" and "unabashedly reductionist" (Kember, 2013).

Implications of the Metaphysical Question

Foundations of Biology Herbert Simon (1990) writes that the fundamental goal of the sciences is to find invariants. Not all invariants, however, are created equal. Some differ in their scope or degree of generality. Some fields have strict, quantitative, general invariants. Meanwhile, others have invariants that can be described as near, partial, or approximate.

A particularly important class of invariants are called 'laws of qualitative structure' (Newell and Simon, 1976). These laws say something about the essential nature of the systems to which they apply, but the statements they make are often qualitative, rather than quantitative, in nature. Though seemingly simple, Simon argues that such laws are among the most important scientific discoveries that we have made.

One, perhaps *the*, foundational law of qualitative structure in biology is the Cell Doctrine. It states that (biochemical) cells are the building blocks of all living organisms. In other words, the cell is the smallest unit which satisfies the minimal structure and function necessary to be called life.

The functionalist position of Strong ALife stands in stark opposition to the Cell Doctrine. In practice, very few ALife practitioners claim to have realized Strong ALife, and so it is not the case that an alternative minimal unit of life has been proposed. Nontheless, there remains an in-principle objection to the Cell Doctrine from the standpoint of Strong ALife. The Cell Doctrine presupposes that life is necessarily biochemical in nature. Meanwhile, Strong ALife argues that the medium of realization is incidental rather than essential. On this view, it is the functional organization of components, not their underlying structural economy, which gives rise to the property of life.

Ethics If we are to take the possibility of Strong ALife seriously, then we must ask ourselves whether and to what degree living artifacts would be worthy of moral consideration. This complicates the already difficult question: in virtue of what are living things worthy of moral consideration?

Even if we reject Strong ALife or abstain from judgement, we must, from a practical perspective, consider the issue of whether there are contexts in which the use of Instrumental ALife might be morally problematic. That the reasons underlying the behavior of complex systems are not always fully understood means that they may behave in unexpected or unpredictable ways. There is then a question about what constitutes an acceptable risk if such systems are are to be deployed in circumstances where human life or critical infrastructure may be at stake. These considerations merit further examination, but such a discussion is beyond the scope of the present paper.

The Methodological Question

We now come to consider the role of ALife systems in the ontology of scientific practice. In particular, we will focus on ALife systems which are software-based, also known as *Soft ALife* (Bedau and ProtoLife, 2006).

Three Perspectives

Most accounts of the role of ALife systems in scientific practice broadly fall into one of three categories, which we shall term *abductionism*, *digital naturalism*, and *simulationism*.

Abductionism treats ALife systems as tools that aid in theory construction and conceptual exploration. This category includes accounts such as that of Dennett (1994), who argues that ALife systems can be construed as a form of "prosthetically controlled thought experiments". In these accounts, ALife systems are primarily thought of as theoretical models, which may be implemented in simulation. These models may serve to prime intuitions and generate hypotheses regarding living systems, which may later be tested by more traditional means. On this view, the modeling and simulation of ALife systems may best be thought of as a theory building paradigm (Diallo et al., 2013).

Digital naturalism views ALife systems as source of experimental data which can be used to settle empirical questions. This category includes Langton's conception of ALife systems as actual instances of life, which may be used in drawing conclusions about living systems. In general, digital naturalists perceive ALife systems as bona fide examples of the phenomena which they attempt to capture. Thomas Ray's *Tierra* (Ray, 1991) is an example of ALife research from the perspective of a digital naturalist.

Simulationism, or *pseudo-empiricism* regards the study of ALife systems as having a sort of empirical 'flavor', but does not perceive such systems as direct sources of experimental data. On this view, ALife systems may be perceived as rich 'artificial worlds' worthy of scientific investigation in their own right (Bullock, 2014). They may provide insights into the dynamics which govern living systems, but do not themselves qualify as experimental systems from which to draw conclusions about the living systems themselves.

Here it will be pertinent to note that one's stance on the metaphysical question has the potential to influence the way in which one engages the methodological question. In the context of Soft ALife, the distinction between Strong ALife and Weak ALife can be viewed as that between computer realization and computer simulation of life. If one subscribes to the Strong ALife point of view, then digital naturalism is likely to be an appealing position. The construction and study of sufficiently advanced ALife systems would, in principle, be ontologically equivalent to traditional experimentation.

Implications of the Methodological Question

The methodological question is intertwined with a number of epistemic concerns regarding ALife research. We are interested in ascertaining character of the knowledge we acquire from building and studying ALife systems. Specifically, we wish to know

1. What is it that we actually learn *about* from pursuing such a course of inquiry?

2. How can we trust that our constructions accurately instantiate or reflect the phenomena they attempt to capture?

3. What justifies the inferences we make from the behavior of these systems to the behavior of living systems at large?

Knowledge About What? For the abductionist, any knowledge gained about an ALife system is strictly theoretical knowledge about the dynamics of a particular model being proposed or examined. Though the ultimate aim of the model may be to describe the behavior of some particular class of living systems, it is the model itself and not the target system which is the object of study.

The digital naturalist, meanwhile, will claim that the output of their investigation is knowledge about a genuine living artifact in particula. They may argue that this knowledge is applicable to some more general class of living systems as well.

Here the simulationist resides in a camp similar, but not identical, to that of the abductionist. Simulationists generally take themselves to be learning about the organization and behavior of objects in the artificial world or virtual laboratory of their construction. These may resemble naturally occuring systems, but the latter are not the targets of their inquiry.

The Problem of Justification A traditional distinction in philosophy of science divides scientific practice into two distinct contexts. The context of discovery, which includes activities related to theory construction, is not thought to governed by any rules. In generating models and hypotheses to account for phenomena of interest, anything is fair game. On the other hand, in the context of justification, which includes activities related to theory testing and experimentation, the scope of legitimate activities is strictly constrained. This is because it is within the context of justification that scientific beliefs are acquired, so actions performed therein must be epistemologically justified.

Under which context does the construction of ALife systems fall? The abductionist is likely to see ALife systems as belonging to the context of discovery. Consequently, there is no immediate problem of justification for the abductionist. The abductionist's model is a theoretical proposal, amenable to vetting and testing at a later stage of inquiry via other means. Insofar as these systems are and remain theoretical objects, the process of their construction is not constrained in any way.

The digital naturalist claims that their systems are bona fide instances of the phenomena they are interested in studying. If this is indeed the case, then conclusions drawn about these systems ought to be true *prima facie*. The very existence of these systems serves as their own justification.

The case of the simulationist is less clear. On the one hand simulationism makes no claim to empricial content with respect to questions about living systems. At the same time, such inquiries are in some sense experiments that revolve around the contents of the simulation. Simulation in general, and ALife in particular, is a deeply troublesome case for the distinction of contexts proposed above. It may be the case that these systems represent a 'third way' of doing science (Ruas et al., 2011), one which does not map neatly onto either side of the traditional divide.

Given some set of conclusions regarding a particular ALife system, how do we go about translating those conclusions some more general class of living systems?

For the abductionist, the answer is straightforward (at least in theory). The predictions of ALife models are to be tested in the same manner as any other proposed theory within a given domain.

The digital naturalist will argue that conclusions drawn about a particular ALife system are equally applicable to some class of living systems (whether possible or actual). Skeptics will argue that the digital naturalist remains responsible for providing philosophical and scientific grounding for their systems.

The simulationist begins with the premise that the knowledge gained from an ALife inquiry is encapsulated within the scope of the simulation. Extending these conclusions to living systems at large requires exiting this framework and adopting an approach similar to that of the digital naturalist.

References

Bedau, M. and ProtoLife, S. (2006). Automated design of artificial life forms from scratch. In *Explorations in the Complexity of Possible Life: Abstracting and Synthesizing the Principles of Living Systems: Proceedings of the 7th German Workshop on Artificial Life, July 26-28, 2006, Jena, Germany*, page 6. IOS Press.

Berryman, S. (2010). Democritus. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2010 edition.

Bullock, S. (2014). Levins and the lure of artificial worlds. *The Monist*.

CaltechArchives (1988). Richard feynman's blackboard at time of his death. File: 1.10-29.jpg.

Darwin, C. (1859). On the origins of species by means of natural selection. *London: Murray*.

Dennett, D. (1994). Artificial life as philosophy. *Artificial Life*, 1(3):291–292.

Dennett, D. C. (1995). Darwin's dangerous idea. *The Sciences*, 35(3):34–40.

Descartes, R. (1985). *The Philosophical Writings of Descartes: Volume 1*. Cambridge University Press.

Diallo, S. Y., Padilla, J. J., Bozkurt, I., and Tolk, A. (2013). Modeling and simulation as a theory building paradigm. In *Ontology, Epistemology, and Teleology for Modeling and Simulation*, pages 193–206. Springer.

Harrison, P. (1992). Descartes on animals. *Philosophical Quarterly*, 42(167):219–227.

Hobbes, T. (1651). *Leviathan*. Dover Publications.

Kember, S. (2013). *Cyberfeminism and artificial life*. Routledge.

Langton, C. G. (1989). Artificial life, santa fe institute studies on the sciences of complexity. In *Proc*, volume 6.

Langton, C. G. (1993). Artificial life. *Appears as*.

Levy, S. (1992). Artificial life: The quest for a new frontier. *New York, New York: Pantheon*.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

Ray, T. S. (1991). An approach to the synthesis of life.

Ruas, T. L., Marietto, M. d. G. B., de Moraes, A. F., Batista, R. d. S. F., Heideker, A., and Aguiar, E. (2011). Modeling artificial life through multi-agent based simulation. *Multi-Agent Systems-Modeling, Control, Programming, Simulations and Applications, Intech*.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, 3(03):417–424.

Simon, H. A. (1990). Invariants of human behavior. *Annual review of psychology*, 41(1):1–20.

Smith, J. (1992). Byte-sized evolution. *Nature*, 355(6363):772–773.

Von Neumann, J. (1951). The general and logical theory of automata. *Cerebral mechanisms in behavior*, pages 1–41.