Biol Philos
DOI 10.1007/s10539-009-9156-z

Synthesizing insight: artificial life as thought experimentation in biology

Liz Stillwaggon Swan

Received: 27 July 2008 / Accepted: 9 February 2009

Abstract What is artificial life? Much has been said about this interesting collection of efforts to artificially simulate and synthesize lifelike behavior and processes, yet we are far from having a robust philosophical understanding of just what Alifers are doing and why it ought to interest philosophers of science, and philosophers of biology in particular. In this paper, I first provide three introductory examples from the particular subset of artificial life I focus on, known as 'soft Alife' (s-Alife), and follow up with a more in-depth review of the Avida program, which serves as my case study of s-Alife. Next, I review three well-known accounts of thought experiments, and then offer my own synthesized account, to make the argument that s-Alife functions as thought experimentation in biology. I draw a comparison between the methodology of the thought-experimental world that yields real-world results, and the s-Alife research that informs our understanding of natural life. I conclude that the insights provided by s-Alife research have the potential to fundamentally alter our understanding of the nature of organic life and thus deserve the attention of both philosophers and natural scientists.

Keywords Artificial life · Thought experiment · Simulation

Introduction: philosophical foundations of artificial life

Technically speaking, artificial life (Alife) is an interdisciplinary field of research incorporating the natural sciences, scientific computing, and robotics, and is characterized by innovative attempts to simulate and synthesize lifelike processes through artificial (in vitro, in silico, or in theorio) means. But what is Alife philosophically speaking? Should philosophers pay it any heed, and why? In a short

L. S. Swan (✉)
Philosophy PostDoc, State University of New York at Buffalo, Buffalo, NY 14261, USA
e-mail: ems52@buffalo.edu; stillwag@mailbox.sc.edu

Published online: 20 February 2009

piece entitled "Artificial life as philosophy," Daniel Dennett urged philosophers not to consider Alife as just another phenomenon in need of critical philosophical analysis but rather as a *new sort of philosophy,* the practice of which consists in the creation of "thought experiments of indefinite complexity" (Dennett 1994). I agree with Dennett's prescription: the collection of research activities that take place under the heading of Alife is best understood as a novel type of thought experimentation—one that stretches the bounds of our common conceptions of natural life.

Traditionally, thought experiments have been most commonly employed in certain areas of philosophy, and in physics, so we are unaccustomed to finding them elsewhere. But, as we will see, Alife provides a wide variety of means for rethinking our conceptions of life and our understanding of evolutionary processes by creating imaginative alternatives to what is—allowing us to entertain what might be, or what could have been, given different parameters—which is essentially how thought experiments work. Fascinating discoveries are being made in Alife about the nature of complex systems, ones that will either force us to accept that we can deepen our understanding of certain features of natural systems through the study of analogous features in artificial systems, or else will challenge us to find grounds for rejecting theoretical extrapolation of this kind. Getting a philosophical handle on how Alife works like thought experimentation will help philosophers orient themselves to this 'new sort of philosophy' and enable them to judge for themselves to what extent, if at all, the insights provided by Alife research will inform traditional philosophical questions regarding, for example, the nature of life, the nature of mind, and the connections between life and mind.

Making this judgment will require a careful look at the work being done in Alife. Four seminal experiments from soft Alife (s-Alife) research are reviewed early on in the paper: three are introductory examples of well-known s-Alife experiments, and the fourth, which I analyze in considerably more depth, reviews the work being done by a team of Alife researchers who created a software platform called Avida to test a specific thesis in Darwinian evolutionary theory. I focus herein only on s-Alife (i.e., that created by scientific computing) because this particular subset of Alife most clearly demonstrates my argument that Alife functions as thought experimentation in biology,[1] and also because it is accompanied by a considerable body of existent literature that is accessible to philosophers not acquainted with the technical side of Alife.[2] Next, I present three well-known accounts of thought experiments and then offer my own which incorporates important features from the other three, and broadens the traditional definition of thought experiments to encompass the thought-experimental nature of s-Alife research. I conclude that because thought

[1] Though I have argued elsewhere (Stillwaggon 2006) that hard Alife, especially Brooks' embodied robotics, provides an attractive antidote to philosophy of mind's traditional thought experiments which make the dubious presumption that it is coherent to talk about disembodied yet conscious experience.

[2] See for example the essays in Margaret Boden's anthology, *The Philosophy of Artificial Life*, which concern the theoretical and epistemological implications of Alife research. For those a bit more technically-minded, I recommend the collection of articles by Rodney Brooks (some of which are coauthored), who is adept at explaining his revolutionary methods in robotics to those lacking a scientific background; see for example Brooks (1991).

experiments have historically been quite informative in philosophy and physics, and because s-Alife works in essentially the same way, any insights it provides likewise deserve philosophical and scientific consideration. By extending a familiar philosophical framework to a new domain of scientific inquiry, I offer non-Alifers a way to understand the work being done in Alife. And by seeing Alife as thought experimentation in biology, we can use to our advantage our familiarity with thought experiments and our knowledge of the various ways of assessing them to make some progress toward developing a robust philosophical account of Alife.

Four different kinds of alife

There are important divisions within Alife research, commonly delineated as wet Alife, hard Alife, and soft Alife. Wet Alife projects are most intimately tied to traditional biological research methods in that their objective is to build hybrid systems comprised of both wet (i.e., natural, biological) and artificial (for instance, silicon) parts. Hard Alife refers to the design and implementation of embodied, artificial creatures, i.e., robots. And s-Alife, on which I focus in this paper, is comprised of computer programs designed either with the practical aim of capturing lifelike behavior artificially, or with the more theoretical aim of testing certain theories in biology that would be very difficult if not impossible to test using natural media. S-Alife comprises much of the work that is of interest to biologists since most of the simulations designed to study characteristic processes of living systems, such as reproduction (or self-reproduction) and adaptive evolution for example, are run using simulated agents in simulated environments.

Though wet, hard, and soft Alife experiments comprise the bulk of the research in the field, there is a fourth, lesser known kind of Alife known as *in theorio* (meaning, in theory, or theoretical), the essence of which is that *theoretical* lifelike behavior is multiply realizable. A famous example of in theorio Alife is the game of *Life*, created by the mathematician John Conway in the late 1960s (Gardner 1970). The game was a physical manifestation of John von Neumann's thought-experimental, self-reproducing automaton, and could variably be 'played' on a checkerboard, a computer using cellular automata, or much more colorfully (and hypothetically), with helicopters hopping to and from the tops of tall buildings. In theorio Alife is concerned with the logic of life, whether it is captured symbolically in mathematics or logic, or theoretically in the imagination. This particular kind of Alife highlights a lurking presumption of the entire field, namely that life may be more abstract than we think, and therefore the actual manifestation of life, whether biological or theoretical, is less important.

The theoretical nature of the research being done in s-Alife raises a host of philosophical questions concerning the epistemological value of experiments in which the agents and the environments are both computer-generated. If the experimental apparatus is simulated, how can it generate results that are *real?* Just because programmers can make computers do things that are fun and interesting, it doesn't mean that real knowledge is being generated, does it? Aren't the experimental results in effect predetermined by the experimenter? These concerns are reminiscent of those regarding the validity of computer simulations in general,[3] and also of thought experiments. But before we can approach the philosophical questions concerning the validity of s-Alife experiments, we need to get clear on exactly how they work, and why they ought to be of interest to philosophers in the first place. This paper is a contribution to that end.

Soft-Alife experiments

In thought experiments the 'action' takes place in the fabricated world. When Galileo imagines himself dropping light and heavy balls from a tower, he discovers that all objects must fall with the same rate of speed regardless of weight since either alternative (the light ball falls faster or the heavy ball falls faster) involves a contradiction. The action in this case, and indeed in all thought experiments, is simulated in one's imagination. The conclusion Galileo derived from the simulated experience supplanted Aristotle's erroneous belief that heavy objects fall more quickly than lighter ones, which had been scientific dogma for almost 2000 years. Thought experiments have the unique virtue of delivering fact from fiction. Provided the thought experiment is a successful one,[4] the insights garnered from the imagined scenario are applicable to the real world. This virtue is arguably shared by s-Alife experiments as well. We will consider some examples.

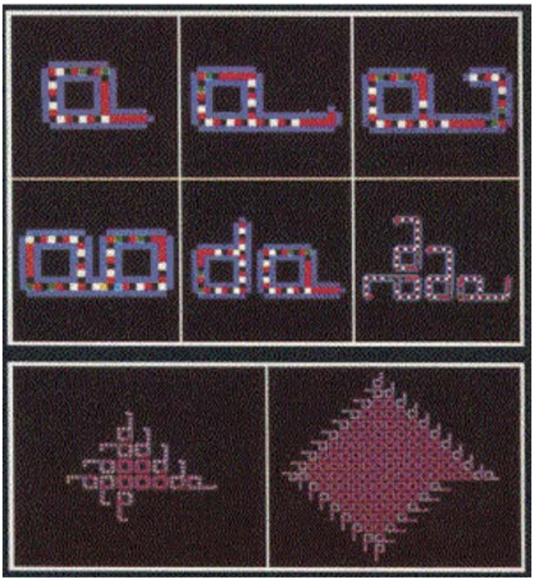
Three introductory examples

(1) Inspired by John von Neumann's thought-experimental self-reproducing automata, Christopher Langton used cellular automata to create "the simplest known structure that can reproduce itself" (Boden 1996). These Langton loops, as they have come to be called, are computational structures shaped like the letter 'Q' and composed of pairs of numbers that are instructed to add more units to the tail each time they reach the corner joining the tail and square, so that eventually an entirely new loop is created (see Fig. 1). The study of Langton loops may help to elucidate the complex relationship between genotype and phenotype in organic organisms—in particular, the question of how a simple set of instructions can both replicate itself *and* produce a new organism, which is difficult to untangle from observing only organic organisms.
(2) Craig Reynolds' Boids (a contraction of 'birdoids') are individual computational organisms, aerodynamically shaped and programmed with three simple rules: (1) stay close to (but do not collide with) other objects in the environment, including other Boids; (2) maintain equal velocities with the Boids in your area; and (3) move toward the perceived center of mass of

[3] See for example: Fox Keller (2003), Hughes (1999), Lenhard (2004), Rohrlich (1991), and Winsberg (2003).

[4] Which on my account (2002) requires that the conditions of the thought experiment be new, relevant, and comprehensive, so that the object of inquiry is mentally manipulated in ways which are novel, reveal something useful, and take into account as many relevant variables as possible.

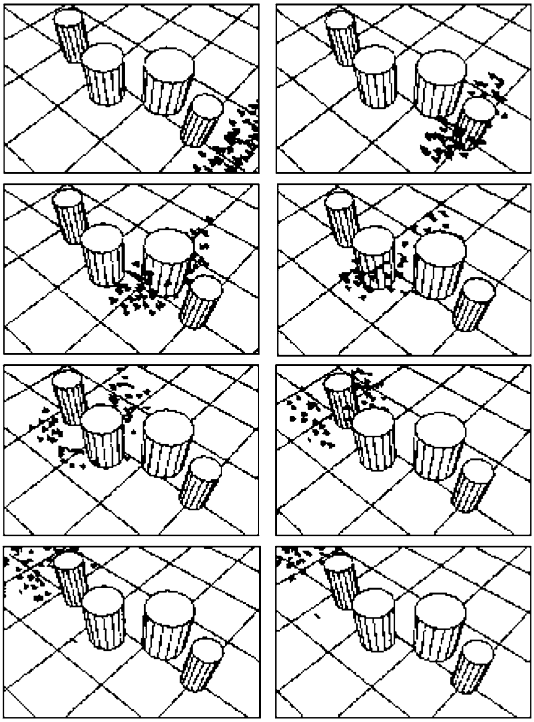
Fig. 1 Langton loops



the Boids in the area. When many Boid programs are run simultaneously, they flock—spontaneously splitting into two flocks to avoid an obstacle and rejoining into a single flock once past the obstacle—though no explicit instructions for flocking behavior have been programmed into the individual boid programs (see Fig. 2). The counterfactual world of simply programmed silicon organisms yields insight into the plausible mechanics of natural bird flocking. And of broader philosophical and scientific interest, Boids provide an example of complex behavior emerging from multiple, simple components, which suggests this phenomenon may be common to both natural and artificial systems (Boden 1996).

(3) An evolutionary biologist named Thomas Ray created the virtual world of Tierra to discover whether the artificial system, initially populated with very simple digital organisms, would eventually exhibit features of natural evolution. Ray's intuition proved correct—Tierra produced many more organisms of greater complexity that interacted in ways Ray interpreted as predator–prey and parasite–host relationships. Ray calls his approach 'synthetic biology' because he is of the Alife camp that believes the creations are not mere models of life, but individual instances of such. He advises that what is important is *"to respect the nature of the medium into which life is being inoculated, and to find the natural form of life in that medium,* without inappropriately trying to make it like organic life" (Ray 1995, italics in original). Tierra can be dualistically conceived both as a singular thought experiment and as a platform for running a variety of individual thought experiments designed to explore particular questions about evolutionary processes.

Fig. 2 Boids in action



Case study: the avida software platform

One persistent challenge to the theory of evolution is to explain how complex features of biological organisms evolved. Darwin's theory on this particular question of evolution is that complex features evolve through the modification of simpler features and functions so that the evolution of complexity is cumulative. Real world constraints such as an imperfect fossil record, incomplete knowledge of the genetic components comprising such features, and the extinction of intermediate forms render infeasible the testing of this theory through traditional methods. Taking a novel approach to an old question, a team of researchers in subspecialties of biology and computer science at Michigan State University designed a software platform called Avida (Lenski et al. 2004) to explore, with the use of digital organisms, the possible conditions under which a specific complex feature could evolve from the modification of simpler features. The researchers define digital organisms as "computer programs that self-replicate, mutate, compete and evolve" (ibid), and differentiate them from computer viruses by pointing out that the latter require direct intervention to mutate and evolve while digital organisms mutate randomly and evolve spontaneously. They borrow Dennett's dictum that evolution will occur "whenever and wherever three conditions are met: replication, variation (mutation), and differential fitness (competition)" (ibid), and explain that their research interests with Avida are focused on uncovering the principles relevant to any evolving system.

The experiments run in Avida were consistent with the basic framework of Darwinian evolution in that competition for resources (energy) plus random mutation comprised the basic mechanism of selection. The initial ancestor organism could replicate but could not perform any of the nine logic functions comprising the experimental set. The digital organisms replicated and mutated while competing for energy which was ‘awarded’ in amounts correlated to the difficulty level of the logic function performed. Many populations of the digital organisms evolved the capacity to perform EQU or ‘equals’, identified by the researchers as the most complex logic function in the set. What they found was that the genotype of the organisms that could perform this function differed from their parents by only one or two mutations, yet differed from the ancestor by many mutations that were apparently crucial to the execution of EQU. This suggests that the ability to perform the most complex function depended indirectly but crucially on each of the prior, simpler functions in that organism’s developmental history. This fact was difficult to see except retrospectively because, as is often the case in biological organisms, initially the mutations were often neutral or even harmful to the organism itself, and only became beneficial when combined with subsequent mutations.

A particularly useful feature of Avida is that it allows for retrospective, counterfactual experimentation. For example, the researchers noted that the genome of the first organism capable of performing EQU had 60 instructions and that eliminating any one of a particular set of 35 of them destroyed that function, so that, “although more than two dozen mutations were used to build EQU, undoing any one of them destroyed this function,” (ibid). It seems there may be a threshold of complexity that is crossed with a single, crucial mutation, rendering the organism capable of performing the complex function. But it is important to note that the initial activation of this capability nevertheless depends in a crucial way on a substrate network of many prior, interacting mutations. Another interesting finding is that EQU is multiply realizable—it can be expressed by several different combinations of instructions, insertions, and deletions, given the right genomic context. So while several simpler functions evolved before EQU in every population, no particular function was essential for the evolution of EQU because each one of these simpler functions was missing from at least one of the parent genotypes. In other words, there was no specific path from simplicity to complexity that was followed each time in the evolution of the EQU function—several different actual paths led to the same ultimate function.

The Avida team believes their findings support the theory that complex features evolved through the gradual modification of existing structures and functions. Furthermore, they believe their research shows the multiple realizability of such complex features. Given the fact that EQU did evolve with high frequency in the experimental populations, coupled with the extreme unlikeliness of there being only one trajectory from simple to complex features, they conclude that there are several possible trajectories. The authors explain that, “Of course, digital organisms differ from organic life in their genetic constitution, metabolic activities and physical environments. However, digital organisms undergo the same processes of reproduction, mutation, inheritance and competition that allow evolution and adaptation by natural selection in organic forms" (ibid). Also, "digital organisms provide opportunities to address important issues in evolutionary biology. They are particularly well suited to problems that are difficult to study with organic forms owing to incomplete information, insufficient time and the impracticality of experiments" (ibid).

In each case discussed above, the fabricated organisms (or organisms plus world) prove insightful in that they are able to mimic organic organisms closely enough to provide genuine instances of the life processes that interest biologists. Some Alifers claim that digital organisms are genuine instances of life (e.g., Ray 1995) while others take a more conservative approach, maintaining that what is important is that these artificial organisms demonstrate complex behaviors that are indistinguishable from those exhibited by organic organisms (e.g., Lenski et al. 2004). Our own philosophical intuitions will dictate whether we take digital organisms to be genuine instances of life or only simulators of life processes. If we choose to count digital organisms as living things, then we have changed the definition of life to include non-biological 'life'. And if instead we argue that digital organisms afford only simulations of life processes, we nonetheless have changed how we think of life, namely by shifting the focus from living *things* to life *processes*. For example, Langton loops constitute a genuine instance of self-reproduction, Reynolds' boids constitute an genuine instance of flocking, and Tierra hosts multiple instances of evolutionary processes. And, crucially, these facts do not change depending on whether we take the agents to be *really alive* or not. Either way you look at it, Alifers are challenging our comfort zone of how we conceptualize life.

What is really going on in each of these examples of synthesized, albeit genuine, instances of evolutionary processes occurring in artificial media? How is it possible that computer programs can be designed to effectively simulate (or synthesize) in digital organisms processes that some biologists claim are indistinguishable from those exhibited by organic, natural organisms? The phrase 'fact from fiction' was used earlier to capture the essence of thought experimental methodology, and this paper makes the argument that the same phenomenon is happening in s-Alife methodology. Thought experiments allow us to imagine in a structured way how things could be in another possible world, or might have been in our actual world given different parameters or different initial conditions. For questions in biology, we have only one reality—the way things *in fact* are given the actual historical contingencies of evolution. But Alife programmers have discovered some very interesting ways to go beyond *life-as-we-know-it* and imagine *life-as-it-could-be*.[5] To understand how s-Alife experiments work like thought experiments, and why this is a good way to understand them, we will consider three well-known accounts of thought experiments plus a synthesized account which I put forth here.

[5] This comparison comes from Christopher Langton, who organized the first Artificial Life workshop in Los Alamos in 1987. His comprehensive manifesto on the research field is reprinted in Boden (1996).

Three thought-experimental frameworks

In this section, I provide brief outlines of three well-known accounts of thought experiments, and then propose my own synthesized account which is consistent in important ways with the other three accounts yet serves to broaden our traditional understanding of thought experiments to encompass the thought-experimental nature of s-Alife research. The relevant comparison to keep in mind here is that between the imagined thought-experimental worlds that yield real-world facts (recall Galileo's Falling Bodies) and the synthesized worlds of digital organisms that provide insight into the nature of organic life. The three accounts of thought experiments I discuss are those of John Norton who claims that thought experiments are nothing more than dressed up arguments, Roy Sorensen who proposes a naturalistic understanding of thought experiments, and Jim Brown who argues for a Platonic understanding of how thought experiments work.

Thought experiments as arguments

Norton's view is, I believe, amenable to the idea that a thought experiment may serve as an intriguing invitation to do a real experiment, or, alternatively, that a thought experiment may provide one with identifiable reasons for redesigning an already run (real) experiment, but his main point is that thought experiments are really just arguments (Norton 2004). On Norton's view, for example, Galileo's Falling Bodies thought experiment is an example of the indirect form of argument known as *reductio ad absurdum.* When Galileo imagines dropping from a tower a light ball attached to a heavy ball, he reasons that on Aristotle's view, the light ball will act as a drag on the combined system (i.e., light ball plus heavy ball) and slow it down. Yet, puzzlingly, the heavy ball will increase the overall weight of the combined system and thus accelerate its fall. Since it is absurd to assert that the combined system falls both more slowly and more quickly than the heavy ball alone, Galileo rightly concludes that the two balls of different weights fall with equal speed since this is the only logical conclusion. As far as this example goes, Norton's view that thought experiments are just arguments seems right on target—Galileo's Falling Bodies clearly is an argument, namely a reductio ad absurdum argument. It should be noted, however, that the example of Falling Bodies is an anomaly among thought experiments, most of which are not as easily categorizable.

Thought experiments as a special kind of experiment

Sorensen's naturalistic view of thought experiments is simply that they *are* experiments, albeit a limiting case of them wherein old information is presented in new dress (Sorensen 1998). Thought experiments should, in his opinion, be used as a research tool along with real experiments and other techniques. Sorensen makes the point that thought experiments often perform the function of retrospective evaluation of real experiments. For this reason his view is similar to (although not entirely consistent with, as he makes clear) Thomas Kuhn's argument that thought experiments serve mainly to correct past conceptual mistakes so that all thought experiments reduce to the uncovering of an inconsistency (Kuhn 1977)—a good example of which, again, is Galileo's Falling Bodies.

Thought experiments as platonic

Brown's position is that thought experiments in the natural sciences are best understood rationally. Thought experiments that function by destroying an old theory and simultaneously erecting a new one, those he calls Platonic thought experiments, work by revealing the laws of nature a priori. He explains that the thought experimenter grasps the laws of nature in the same way the mathematician intuits mathematical truths, i.e., "by a kind of perception of the relevant laws of nature" (Brown 1991). On this account, thought experiments and mathematical objects are tools that allow us to perceive abstract, Platonic, entities.

Thought experiments as abduction: where the a priori and the empirical converge

My own view is that the Avida software, and s-Alife experiments in general, comprise a special type of *abduction*, the third kind of logical inference first identified by Charles Sanders Peirce in the 1870s (Peirce 1958). Abduction has been defined variously by Peirce and others as: the faculty of our reason responsible for hypothesis-formation and hypothesis-correction, perceptual knowledge, common sense beliefs, and inference to the best explanation (and there are probably others). Peirce believed that abduction worked in tandem with deduction and induction, and he developed an account of how the three interact in scientific practice. The feature of abduction that makes it the best candidate for what s-Alifers are doing is that it allows for the possibility of insight without evidence (which amounts to a priori reasoning). Abduction allows one to *reason from effects back to probable causes* even when direct observation of such causes is strictly impossible, as is the case, for example, in testing various aspects of evolutionary theories.

This epistemic feature of allowing one to reason from effects to (probable) causes is especially salient in the Avida research as experimenters are able to directly manipulate individual genes in the genotypes of digital organisms to test whether the gene's removal (or addition) will produce any effect on the ultimate EQU function. Individual genes may seem harmful when they first appear, but once the EQU function has been acquired, one can experiment backwards as it were and discover that each of the intermediary genes was in fact necessary for the development of the most complex function, EQU. Similarly, in Langton Loop experiments, we do not yet fully understand the intricacies of the relationships between genotype and phenotype in organic organisms, and it is believed that by observing and manipulating the mechanisms that synthesize this relationship artificially, we will gain insight into the analogous mechanisms in the natural case. In sum, the line of reasoning is from artificial effects to natural causes. And likewise with 'boids': if three simple rules are sufficient to effect lifelike flocking in boids, this suggests that similar mechanisms are at work in natural bird flocking. Again, the reasoning is from known effects to probable causes.

Abduction as a synthesized account of thought experiments

Conceptualizing thought experiments as abductive allows for the inclusion of the important features of the three well-known accounts outlined above, as well as the expansion of our definition of thought experiments to encompass the creative, and sometimes retrospective reasoning characteristic of s-Alife experimentation. Because on Peirce's account abduction works in tandem with deduction and induction, conceptualizing thought experiments, and thus s-Alife experiments, as abductive allows for Norton's insight that thought experiments are a special kind of argument. Though s-Alife programs are not strictly arguments since lines of code are commands rather than assertions that can be assigned truth-values, the methodology of s-Alife experimentation, which includes the process of hypothesis-formation, is abductive and thus can be thought of as a special kind of argument-formation.

Sorensen's insight that thought experiments are just a particular kind of experiment is useful too. S-Alife experiments go beyond the uncovering of contradictions or conceptual mistakes (which Sorensen claims is the main function of thought experiments), but they clearly do comprise a special kind of experimentation. Although the Avida research is abductive in the sense of providing insight into a theory of Darwinian evolution without direct evidence, the Avida methodology is clearly empirical too. In thought experiments and Alife experiments, the empirical and the a priori feed off one another. In fact, Alife research as a whole can be conceptualized as a collection of methods employed to empirically verify the a priori premise that our current concept of life is in need of revision. One could argue that the Avida team had intuited a priori how complex features evolved in biological organisms (and one could make this same argument for Darwin). But it seems that even given an intuitive basis for the creation of Avida, the actual work can only be conceived as empirical, given the generation of the numerous populations of digital organisms and the retrospective experiments run to determine the significance played by each intermediary mutation in the acquisition of the final complex function. Alife provides an excellent example of the fact that thought experiments and scientific experiments interact in interesting ways: either can serve to correct for previous conceptual or empirical errors, and a thought experiment may generate knowledge that is conceptual only until empirically verified at some later time.[6]

Brown's insight that thought experiments are Platonic and thus provide a priori insight into their subject matter captures quite nicely the essence of the Alife endeavor (mentioned above) which is described by its founder, Christopher

[6] Lenski et al's 2008 constitutes an excellent example of how scientific experimentation and thought experimentation can work in tandem. The team used strains of *E. coli* to trace the evolution of a citrate-using (Cit+) ability in the organisms and discovered, by "replaying" evolution from different points in each strain's development, that historical contingencies were highly influential in the ultimate evolution of Cit+. This empirical work with organic media nicely reflects the team's experimental conclusions from their earlier research with Avida; in particular the fact that although there were several possible trajectories in the ultimate development of the EQU function, there were nevertheless specific genes that had to be in place for its development to occur at all.

Langton, as follows: "we expect the synthetic approach to lead us not only to, but quite often *beyond*, known biological phenomena: beyond *life-as-we-know-it* into the realm of *life-as-it-could-be"* (Boden 1996). It could be argued that Alife experiments afford us a priori glimpses into "life as it could be," and that Alifers would not have been able to design appropriate experiments to challenge the accepted definition of life unless they had first perceived that life may be more abstract than we tend to think, i.e., that some or all of the processes we commonly think of as being unique to living systems might also be instantiated by artificial systems. S-Alife experimentation and philosophical thought experimentation are intimately linked in their ability to free the experimenter from real-world constraints—such as incomplete information, or insufficient time—allowing them not just to see what is the case, but also to imagine what could be the case. Alife research is empirical and philosophical thought experimentation works in tandem with the empirical. Most significantly, the methodology is the same—Alifers have established various means for fabricating hypothetical worlds that are populated with hypothetical agents from which they can deduce conclusions that have real-world validity.

Conclusion: What can alife tell us about real life?

Mark Bedau notes that today the question of the nature of life is shunned by philosophers for seeming too scientific and likewise is shunned by scientists for seeming too philosophical (Bedau 2004). Fortunately, Alife has found a home for itself in this neglected conceptual space. The various disciplines interested in Alife are united by the concern that biology is limited for having only earthly (wet and carbon-based) life to assess, and by the intuition that there could be more to life than we have so far assumed. Alifers are challenging our common conceptions of life by posing questions whose answers would be crucial to a comprehensive understanding of the nature of life, yet ones that biologists themselves typically do not ask. For example, many s-Alife experiments, including all of the ones we looked at, question whether life is a process that is necessarily instantiated by wet and carbonic organisms or whether it could be a process that is characteristic of a certain class of complex systems, whether natural or artificial. We have available only one example of life—earthly life—but Alifers have cultivated a program of counterfactual speculation about how life might have evolved differently given different initial conditions (see footnote 5), or how the current concept of life might extend to more phenomena than we have so far assumed. "It could turn out," writes Bedau, "that the fundamental process that produces the familiar phenomena of life is essentially the same process that produces phenomena that we do not today recognize to involve life," (Bedau 2004). It goes without saying that such a discovery would be important to artificial and natural scientists alike.

Alifers are naturally drawn to the notion of life-as-process instead of life-as-living-things because adopting the former stance frees one up to explore the possibility of processes we think of as 'natural' inhering in systems we do not commonly think of in that way. In other words, though we do not think of today's crude robots (by sci-fi standards) as alive in any robust sense, our intuitions would be challenged by a robot that looked, moved, and conversed just like a human. 'Aliveness' would depend in this hypothetical case not on material composition but on the robot's behavioral patterns, which are processes. Defining life in terms of living things, on the other hand, limits one to cataloguing the familiar world around us. If we define life in terms of things that are composed of biological material, or worse, in terms of things that have a special essence (e.g., an *élan vital*[7]), then we cut ourselves off from the *possibility* of asking whether the phenomenon of life may be more abstract than we have come to assume. The above quotation from Bedau exemplifies the fundamental Alife premise that a genuine understanding of life will come not from studying living *things*, but living *processes.* And it also reveals the astuteness of Dennett's characterization of Alife as a *method* rather than a *phenomenon.*

Alife as 'theoretical biology'

Why is Alife sometimes referred to as 'theoretical biology' or 'biology of the possible'? While our intuitions generally serve us well in distinguishing between living and non-living stuff in the world (though the borderline cases such as viruses persistently blur this boundary), defining life in any thoroughgoing way is something that has remained intractable for both biologists and philosophers. Most working definitions of life are actually lists of criteria, characteristics that ostensibly apply to all living things, yet there is good reason to believe that this approach to understanding life may be inherently problematic. Claus Emmeche makes the important point that such criteria change according to the list-maker's outlook—there are physiological, metabolic, genetic, biochemical, and thermodynamic definitions of life, and each definition is limited for identifying purportedly necessary yet not (collectively) sufficient conditions for life. For example, Emmeche explains how the metabolic definition fails to define life uniquely because it states that "a living system is one that is distinct from its external environment and that exchanges material with its surroundings […] without changing its general properties" which, he points out, also accurately describes a whirlpool in a river (Emmeche 1994). The other definitions suffer the same fate of applying not only to living systems such as organisms but also to complex systems that are not living. Traditional approaches to defining life are clearly problematic and cannot boast much success. One of the attractive features of Alife is that it provides a completely novel approach to these outstanding questions, as we saw in the classic examples of s-Alife experiments.

Many s-Alife simulations, as we have seen, are designed to supplement our knowledge of natural systems, specifically those that concern questions of emergence, organization and self-organization, reproduction, evolution and co-evolution, all of which are features of natural systems that are of interest to biologists. S-Alife simulations offer biologists an alternative method by which to study features of natural systems that are difficult to study using traditional methods

[7] Bergson 1911.

for a whole host of practical reasons, one example of which is the time constraint inherent in reproductive studies of natural organisms. Simulated organisms by contrast can be designed to reproduce much more quickly, as we saw in the Tierra and Avida experiments.

Alife provides a means for expanding our definition of life or even fundamentally changing the way we look at life through its recreation of lifelike processes in systems and objects, i.e., programs and robots, that are typically considered nonliving. A clear connection between the methodology of Alife and that of thought experiments can now be made: Alifers are in the business of fabricating alternate life forms and life habitats that are strikingly different from ours yet nevertheless exhibit complex processes and behaviors we would previously have expected to find only within the confines of biological life. Similarly, a thought experimenter postulates other possible realities that are like ours in all relevant ways yet different in one key aspect in order to gain a deeper insight into the actual world we inhabit. The line of inquiry employed in exploring these questions is inherently counterfactual, i.e., asking *what if* x, y, and z were the case? The most appropriate catch phrase to describe the spirit of what Alifers are doing is ‘thinking outside the box’. And now that we have a more robust philosophical understanding of just what Alifers are doing, we are closer to being able to judge to what extent their experimental insights will inform our outstanding philosophical questions on the nature of life processes.

Acknowledgment I would like to extend my sincere thanks to the anonymous reviewer who put almost as much work into this paper as I did.

References

Bedau M (1996) The nature of life. In: Boden M (ed) The philosophy of artificial life. Oxford University Press, Oxford, pp 332–357

Bedau M (2004) How to understand the question ‘what is life? In: Bedau M, Husbands P, Hutton T, Kumar S, Suzuki H (eds) Workshop and tutorial proceedings, ninth international conference on the simulation and synthesis of living systems (Alife IX). Boston, MA, pp 126–129

Bergson H (1911) Creative evolution. Tr., Arthur Mitchell (New York: Dover, 1998 [1911])

Boden M (1996) The philosophy of artificial life. Oxford university press, Oxford

Brooks R (1991) Intelligence without representation. In: Haugeland J (ed) Mind design II. The MIT Press, Cambridge, pp 395–420

Brown JR (1991) The laboratory of the mind. Routledge, New York

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

Emmeche C (1994) Garden in the machine. Princeton University Press, Princeton

Fox Keller E (2003) Models, simulation, and computer experiments. In: Radder H (ed) The philosophy of scientific experimentation. University of Pittsburgh Press, Pittsburgh

Gardner M (1970) Mathematical games: the fantastic combinations of john conway’s new solitaire game life. Sci Am 223:120–123

Hughes RIG (1999) The ising model, computer simulation, and universal physics. In: Morgan M, Morrison M (eds) Models as mediators. Cambridge University Press, Cambridge

Kuhn TS (1977) The essential tension. University of Chicago Press, Chicago

Langton C (1996) Artificial life. In: Boden M (ed) The philosophy of artificial life. Oxford University Press, Oxford, pp 39–94

Lenhard J (2004) Nanoscience and the janus-faced character of simulations. In: Baird D, Nordmann A, Schummer J (eds) Discovering the nanoscale. IOS Press, Amsterdam

Lenski RE, Ofria C, Pennock RT, Adami C (2004) The evolutionary origin of complex features. Nature 423:139–144

Norton J (2004) Why thought experiments do not transcend empiricism. In: Hitchcock C (ed) Contemporary debates in philosophy of science. Blackwell, Oxford, pp 44–66

Peirce CS (1958) Collected papers. In: Hartshorne C, Weib P (eds) Band I–VI. Harvard University Press, Boston, pp 1931–1935 Band VII, VIII. Burks AW (ed)

Ray T (1995) An evolutionary approach to synthetic biology: zen and the art of creating life. In: Langton C (ed) Artificial life: an overview. The MIT Press, Cambridge, p 196

Reynolds C (1987) Flocks, herds, and schools: a distributed behavioral model, in computer graphics. In: SIGGRAPH 1987 conference proceedings, 21(4): 25–34

Rohrlich F (1991) Computer simulations in the physical sciences, PSA 1990, vol 2. The Philosophy of Science Association, East Lansing, pp 507–518

Sorensen RA (1998) Thought experiments. Oxford University Press, London

Stillwaggon L (2002) Delivering fact from fiction: the epistemological value of scientific thought experiments. Philosophy Masters Thesis, San Francisco State University, unpublished

Stillwaggon L (2006) How does alife inform the mind-body problem? In: Artificial intelligence and the simulation of behavior quarterly, vol 123. Winter/Spring, p 8

Winsberg E (2003) Simulated experiments: methodology for a virtual world. Philos Sci 70:105–125