Artificial Biology

C.C. Maley
The Artificial Intelligence Laboratory
MIT NE43-803
545 Technology Sq.
Cambridge, MA 02139
email: cmaley@ai.mit.edu

May 30, 1995

Abstract

Artificial Life spans a continuum from engineering to biology. Biological organisms are inspiring both as existence proofs for the tractability of many engineering problems and as testaments to the innovation and power of evolution as a sort of "blind" engineer. The middle of the continuum is occupied by researchers trying to synthesize life in novel substrates. Far from the nightmares of Frankenstein, theoretical biologists are beginning to open up some of the most intractable problems of biology through the use of new computer modeling techniques. Artificial Biology is this approach of using computer models to simulate biological systems. These simulations allow us to gain experimental control over biological systems that are beyond direct empirical methodologies. The models also allow us to address systems with complexities that would make an analytical model intractable. In this way, Artificial Biology complements the traditional fields of empirical and theoretical biology.

1 The Structure of Artificial Life

We have only one example of how life can work. All of life, as we know it, is based on a genetic code of four nucleic acids and their translation into proteins. Yet, we have begun to suspect that perhaps "life" is a more general phenomenon and so might be found in unfamiliar substrates. It is almost as if we were stuck in the middle ages when our only experience of light was that it came from fire. It is difficult to understand any complex phenomenon when you cannot abstract the concept from its implementation. This is the *raison d'etre* for the field of Artificial Life.[13, 15, 14, 3, 23] In the long view, researchers in Artificial Life seek to simulate and even synthesize life in novel media in order to understand the phenomenon of life itself. Only then can we hope to separate the historical accidents that led to life on Earth from the systematic constraints that shaped it.

Artificial Life can be understood as the froth that is thrown up when the waves of engineering hubris crash upon the shores of organismal biology. It is the intersection of those disparate disciplines. Artificial Life is also an extremely young field, and so it is still poorly defined.

The typical description of the field of Artificial Life divides the field by the medium or substrate of the research. "Hardware, Software, and Wetware." The later is not software for the brain, a la William Gibson, but rather chemical systems that demonstrate life-like properties. Or in other words, Artificial Life in test tubes. Unfortunately, the organization of the field of Artificial Life by substrate obscures the important distinctions between different forms of Artificial Life research. It is perhaps more illuminating to describe Artificial Life research as spanning a continuum from engineering to biology.

2 Engineering from Biology

One extreme of Artificial Life is populated by engineers intent on raiding the secrets of Nature for its insights. Every day of our lives we live amongst and interact with biological machines vastly more complex than anything humans have ever built. In a sense, evolution has "solved" many of the problems that engineers face today. This observation has drawn engineers into an attempt to harness the forces of evolution and thereby sow the seeds of our own designs.

John Holland[10, 8] is credited for inventing genetic algorithms, a bare-bones simulation of natural selection that can be used to solve any arbitrary problem. Genetic algorithms simulate three important aspects of evolution in order to solve a problem. First, they include a population of individuals. Second, these individuals vary by how successful they are at surviving and reproducing. Finally, classic genetic algorithms include the processes of recombination and mutation which inject new variation into the next generation. An individual in a genetic algorithm is usually a bit string encoding of a potential solution to the problem. The relative quality of these potential solutions are measured by a "fitness function" which maps the domain of solutions into a range of scores. These scores are then used to preferentially allow the better solutions/individuals to survive and reproduce. This results in the creation of the next generation of potential solutions with the addition of a little mutation. The process can be repeated over many generations until an adequate solution is discovered. Holland has come to be known as "the man who taught computers how to have sex."

Other engineers, working primarily in autonomous robotics, find inspiration in the neurophysiology of animals, as well as the mechanics of their anatomy.[17, 19, 18, 5] But regardless of whether engineers are looking at the scale of evolving lineages or the scale of neurons in an insect, they are using biological examples as existence proofs of the fact that the challenges of engineering can be overcome.

3 Synthetic Life

Somewhere between engineering and biology, amidst the froth, lie the people who are trying to create new forms of life, that is, artificial or synthetic life. In a reference to the old debate in Artificial Intelligence, these people are sometimes called "strong artificial life" researchers. To be fair, few people identify themselves as such. However, the prospect of actually creating new forms of life lies at the heart of Artificial Life and sets our hearts a-flutter. If we ever could create new forms of life, we would no longer have a sample of one to study. Comparative biology, in a deep sense, could begin.

Tom Ray's Tierra model[20, 21] is the best known example of the synthetic life initiative. Tierra is similar to game of "Core Wars" where little assembly language viruses live, replicate, and destroy each other in a section of "core" memory. However, Tierra prevents the self-replicating programs from destroying each other. Ray also forces low levels of mutation into the process of reproduction. Over time, these "organisms" evolve to interact and exploit each other in the constant struggle for memory space and processing time. The result is a digital ecosystem of sorts. Are these self-replicating "organisms" alive?

Lacking a definition of life, it is difficult to determine when we have actually managed to synthesize a new form of life. For the moment we are stuck relying on our intuitions. However, our intuitions have been formed through our experience with only a single example of life. The end result may well be a moving goal post. Given this difficulty, researchers in synthetic life might do well to directly address specific definitions of life. At the very least, we could use the results of their endeavors to refine our intuitions about the nature of life.

4 Artificial Biology

On the far end of the continuum, opposite the engineers, are the biologists trying to investigate the intractable problems of biology through computer modeling.

Given our single example of a living system, we might still be able to tease apart systematic constraints from historical accidents if we were able to "run" many "trials" of evolution under identical initial conditions. Unfortunately, the time scales and complexity of living systems make such an experiment extremely difficult, if not impossible. Traditional approaches in biology include studying very short life cycle organisms in the lab and identifying "observational experiments" in the fossil record. That is, it may be possible to study the evolution of different fossil lineages under similar environments. However, the lack of experimental control and the difficulty of extracting detailed information from the fossil record severely limit this approach. Artificial Biology supplements these approaches by modeling biological systems. Thus we may gain a high degree of experimental control at the expense of a modeling abstraction.

A model can helps us to distinguish systematic from stochastic phenomena in biological systems. Varying only the initial random number generator seed in a model can provide different coincidental effects starting from the same initial conditions. Experiments can also be carried out through the systematic modulation of the parameters of the model so as to examine the effects of different constraints on the evolution of the system. The model thus acts as a sort of crystal ball. It gives us a glimpse of what might have been had fate taken a different path, and it also gives us a glimpse of what could be, given different constraints on the biological system.

Artificial Biology is, in effect, an extension of theoretical biology. Computer simulations provide a new tool for developing biological insight. Configuration models[4], where each individual and the rules for their interactions are explicitly represented in the computer, complement the analytical models of traditional theoretical biology.

4.1 Methodology

Progress is made in Artificial Biology by examining hypotheses in biology and determining how well they apply to a synthetic system. The methodology follows in four steps.

1. Isolate a question.
2. Identify the minimal requirements.
3. Build and Run a model from those requirements.
4. Either:
   (a) Revise the question, or
   (b) Accept the results predicated on the assumptions.

First a question must be isolated and well defined, so that it can be translated into a model. In order to make this translation, the minimal components necessary for testing the hypothesis must be specified. Those minimal requirements can then be built into a model. Once the model has been used to collect data, the results can be treated in one of two ways. First, it is quite possible that the minimal components necessary to test the hypothesis will lead to such a simplified model, that the results will appear biologically implausible. In this case, we have learned something about the hypothesis. The hypothesis must rest on further assumptions or elaborations. Once those have been identified, we can start the modeling process over again. When the model reflects a sufficiently complete hypothesis, the results of the model can be used to either reject or support the hypothesis.

This methodology has the benefit of providing useful results regardless of the biological plausibility of the results. We are able to refine our biological hypotheses through the process of modeling. On the one hand, the model may elucidate unspoken assumptions behind the hypothesis. The very process of constructing a model forces us to spell out all of the relationships between organisms and the environment. Frequently, this endeavor illuminates large gaps in our knowledge of the system. While the artificial life researcher is forced to make some minimal assumptions to fill in those gaps, field and laboratory biologists can then test those assumptions. In this way, we complete the cycle between theoretical and experimental biology.

4.2 Sexual Selection

Collins and Jefferson[6] provide a good example of an Artificial Biology model. They chose to examine the dynamics of sexual selection. Sexual selection is one of the five major processes that leads to evolution, along with natural selection, mutation, migration, and genetic drift. While we tend to focus on the struggle for survival, the domain of natural selection, the question of who reproduces is just as important as who survives, since an organism must do both in order to propagate its genes into the next generation. The selection of a mate is an important dynamic in the process of reproduction for many species. This selection can be based on characteristics that have nothing to do with the mate's ability to survive. In fact, sexual selection is thought to be the driving force behind the evolution of the gaudy, but presumably burdensome, peacock's tail.

How does the phenomenon of sexual selection evolve in a population? Collins and Jefferson began with a two loci, two allele analytical model developed by Kirkpatrick[12]. One locus represented a gene for some trait to be expressed in males and the other a gene expressed in females for the preference for that trait in their mates. Kirkpatrick had shown that there is an infinite set of possible equilibria for the proportion of genes for sexual selection in a population. These equilibria all lie along a curve. The particular point on that curve that will be approached by a population depends on the strength of the mating preference gene and the deleterious effect of the trait on the survival of the males.

Collins and Jefferson examined the dynamics of the model as they relaxed the simplifying assumptions that had been necessary in order to make the analytical model mathematically tractable. They elaborated the model by using finite, rather than infinite, populations, females with only a sample of the males for mate choice, rather than the entire male population, and they also introduced a low rate of mutation. They concluded the initial analysis with:

> ...we have not only verified that the simulation model produces the equilibrium derived by Kirkpatrick, but we have also demonstrated that it is a viable way to study the dynamics of populations far from equilibrium.

More interestingly, Collins and Jefferson continued to extend the model by adding a "stepping stone" demic structure to the population. The population was divided into subpopulations, or demes. An individual was equally likely to mate with any individual in the same deme, but there was relatively little migration between demes. They also examined the population when individuals were allowed to be diploid rather than haploid. That is, an individual was given two values, or alleles, for each gene. This allows alleles to be recessive or dominant. The change from panmixia, where all individuals in the population were equally likely to encounter and potentially mate with all other individuals in the population, to a demic structure does not fundamentally change the equilibrium dynamics of the model. However, the addition of diploidy to a panmictic population dramatically alters the behavior of the population far from equilibrium. The diploid population tends to resist the invasion of new alleles. These insights were opaque to the analytical model.

4.3 Strengths

The Collins and Jefferson model clearly illustrates the ability of a configuration model to fill in a gap between analytical models and empirical methodologies. Many of the dynamics of a biological system can be directly translated into a configuration model. This makes the interpretation, and criticism, of the models relatively easy when compared to the layers of abstraction necessary to build an analytical model. Finally, building a configuration model forces the explicit representation of the assumptions that underlie our understanding of the biological system. These assumptions may then be used to generate research questions for empirical biologists.

4.4 Weaknesses

Artificial Biology can replace neither analytical theory nor empirical research in biology. It should be seen as a complementary paradigm. The weaknesses of Artificial Biology can be understood in contrast to the strengths of these two alternatives. Configuration models are in essence mathematical models, for they are specified in a formal computer language. However, due to their complexity, it is extremely difficult to directly translate a configuration model into an analytic model. And so we cannot bring to bear the powerful tools of mathematical analysis on the problems. We are left to resort to sampling and statistical tools, like empirical biologists, in order to glean knowledge from our models. On the other hand, Artificial Biology does not work with real biological systems, but rather with simulated systems. Any phenomena illuminated through these models must eventually be supported by results from complementary fields of biology before we can say that we have learned anything new about life on Earth.

4.5 The Future of Artificial Biology

For the moment, Artificial Life, and more specifically, Artificial Biology, has opened up vast areas of biology. This is a revolution driven by technology. Only recently have we had computers fast enough to model entire populations of individuals evolving over geological time scales. With these techniques we may finally examine how the microevolutionary processes of natural selection, mutation, sexual selection, migration, and genetic drift may explain our observations of macroevolutionary phenomena like speciation, extinction events, and evolutionary innovations. Many Artificial Life researchers have already weighed into the debate over the progress of evolution. It is a common observation[11, 16, 2, 1, for example] that simulated evolving systems tend to alternate between long periods of stasis and short bursts of change, so called punctuated equilibria.[9, 7]

One of the grand challenges that beckons Artificial Biology is to explain the major transitions in evolution.[22] These transitions all involved the transformation of populations of simple individuals into populations of conglomerates of individuals reproducing as coherent units. At the lowest level, there was a transition from a pre-biotic soup of molecules into self-replicating macro-molecules. Three billion years, and a number of important transitions later, multicellular creatures emerged from a world populated exclusively by single celled organisms. And in the last instant of geological time, human culture has erupted. Replicating information has liberated itself from the mortal coils of genes and now infests patterns of neuronal activation, electronic bits, and the even the squiggles of ink that you have been reading.

References

[1] S. Bankes. Exploring the foundations of artificial societies: Experiments in evolving solutions to iterated n-player prisoner's dilemma. In R. Brooks and P. Maes, editors, *Artificial Life IV*, pages 337–342, Cambridge, MA, 1994. MIT Press. for example.

[2] J. Batali and P. Kitcher. Evolutionary dynamics of altruistic behavior in optional and compulsory versions of the iterated prisoner's dilemma. In R. Brooks and P. Maes, editors, *Artificial Life IV*, pages 343–348, Cambridge, MA, 1994. MIT Press.

[3] R. Brooks and P. Maes, editors. *Artificial Life IV*, Cambridge, MA, 1994. MIT Press.

[4] H. Caswell and A. M. John. From the individual to the population in demographic models. In D. Deangelis and L. Gross, editors, *Individual-based Models and Approaches in Ecology*, pages 36–61, New York, 1992. Chapman and Hill.

[5] D. Cliff, P. Husbands, J.-A. Meyer, and S. W. Wilson, editors. *From Animals to Animats. Proceedings of the 3rd International Conference on Simulation of Adaptive Behavior.*, Cambridge, MA, 1994. MIT Press.

[6] R. Collins and D. Jefferson. The evolution of sexual selection and female choice. In F. J. Varela and P. Bourgine, editors, *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, pages 327–336. MIT Press, 1992.

[7] N. Eldredge. *Time Frames: The Rethinking of Darwinian Evolution and the Theory of Punctuated Equilibria.* Simon and Schuster, New York, NY, 1985.

[8] D. E. Goldberg. *Genetic Algorithms: In Search, Optimization and Machine Learning.* Addison-Wesley, Reading, MA, 1989.

[9] S. J. Gould and N. Eldredge. Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, 3:115–151, 1977.

[10] J. H. Holland. *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor, MI, 1975.

[11] A. R. Johnson. Evolution of a size-structured, predator-prey community. In C. G. Langton, editor, *Artificial Life III*, pages 105–129, Reading, MA, 1994. Addison-Wesley.

[12] M. Kirkpatrick. Sexual selection and the evolution of female choice. *Evolution*, 36(1):1–12, 1982.

[13] C. G. Langton, editor. *Artificial Life*, Reading, MA, 1989. Addison-Wesley.

[14] C. G. Langton, editor. *Artificial Life III*, Reading, MA, 1994. Addison-Wesley.

[15] C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen, editors. *Artificial Life II*, Reading, MA, 1992. Addison-Wesley.

[16] K. Lindgren and M. G. Nordahl. Artificial food webs. In C. G. Langton, editor, *Artificial Life III*, pages 73–104, Reading, MA, 1994. Addison-Wesley.

[17] P. Maes, editor. *Designing Autonomous Agents.* MIT Press, Cambridge, MA, 1990.

[18] J.-A. Meyer, H. L.Roitblat, and S. W. Wilson, editors. *From Animals to Animats. Proceedings of the 2nd International Conference on Simulation of Adaptive Behavior.*, Cambridge, MA, 1993. MIT Press.

[19] J.-A. Meyer and S. W. Wilson, editors. *From Animals to Animats. Proceedings of the 1st International Conference on Simulation of Adaptive Behavior.*, Cambridge, MA, 1991. MIT Press.

[20] T. S. Ray. An approach to the synthesis of life. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 371–408, Reading, MA, 1992. Addison-Wesley.

[21] T. S. Ray. An evolutionary approach to synthetic biology. *Artificial Life*, 1:179–210, 1994.

[22] J. Maynard Smith and E. Szathmáry. *The Major Transitions in Evolution.* Oxford University Press, Oxford, UK, 1995.

[23] F. J. Varela and P. Bourgine, editors. *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, Cambridge, MA, 1992. MIT Press.