1996

Artificial life : life form, simulation, or simulacrum

John Paul Sullins
San Jose State University

ARTIFICIAL LIFE:

Life Form, Simulation, or Simulacrum

A Thesis

Presented to

The Faculty of the Department of

Philosophy

San Jose State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

By

John Paul Sullins III

August, 1996

UMI Number: 1381447

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

APPROVED FOR THE DEPARTMENT

OF PHILOSOPHY

Richard Tieszen

Dr. Richard Tieszen

S N Cook

Dr. Scott David Noam Cook

Rudy Rucker

Dr. Rudy Rucker

APPROVED FOR THE UNIVERSITY

M. Lou Lewandowski

ABSTRACT

ARTIFICIAL LIFE:
Lifeform, Simulation, or Simulacrum

by John Paul Sullins III

This thesis addresses the philosophical implications the emerging science called artificial life (or A-Life). It examines aspects of the role artificial life programming will play in the future of the philosophy of biology, the philosophy of science, the philosophy of technology, the philosophy of computer science and cognitive science. In addition, some key claims made by artificial life researchers, regarding the ability to create living organisms in the medium of the computer are critically examined.

We find that while A-Life projects do fit some of the traditional definitions for living systems, we are still able to cast doubt on the claim that all forms of A-Life can be said to be examples of living things. Specifically we find that A-Life systems that are entirely resident within a computer, such as those based on cellular automata, are particularly vulnerable to criticism. We find that the study of A-Life can be seen as one of the first postmodern sciences in that it is a science that is concerned more with the computational media used to study the phenomena of life, then with actual biological instances of life. The understanding that A-Life studies a simulacrum of life and not life itself is crucial to correctly interpreting the results and findings of this new field of study.

Acknowledgments

This work is dedicated to my Fiancé Nicolet Walker, and to my parents and family.

I would also like to thank the faculty and staff of the philosophy department at San Jose State University for their help and support. And I would like to thank the Dierks-Morgan Scholarship, and the Alumni Association for the many semesters of financial aid given me during the writing of this thesis.

I would like to extend a special thanks to the members of my theses committee for their criticisms and encouragements. No work of this type can be accomplished without the aid of others and I am sincerely grateful to those who have helped me accomplish this goal.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Artificial life (hereafter referred to as A-Life) is the name given to an emerging brand of science. This new science is an interdisciplinary mixture of computer science, theoretical biology, physics, anthropology, and philosophy. More specifically, A-Life is a particular offshoot of a new branch of physics and computer science known officially as "complex systems theory," and popularly as the science of chaos and complexity (Bedau, 1992, p. 494). Simply put, A-Life is the study of man-made systems which exhibit life-like behavior and/or can be claimed to be examples of actual living systems. This definition is imprecise and needs to be expanded but it serves well as a brief sketch of the general concerns of this science.

This thesis will attempt to briefly describe the emergence of artificial life from its inception in the theories which were advanced by John von Neumann regarding self-replicating automata, and then we will look at its current expressions in the field of computer science and theoretical biology. The potential uses of A-Life techniques and experiments in the fields of theoretical biology and computer programming will be outlined. Additionally, this thesis will look at proposals made by prominent philosophers who suggest that A-Life techniques be used to study problems in the areas of Philosophy of Biology, Philosophy of Science, Philosophy of Mind, and Metaphysics. The bearing Artificial Life has on social, political, economic, and ethical philosophies will also be discussed.

The main body of this thesis will be devoted to a discussion of the *strong* claims made by some A-Life researchers which suggest that the theories developed by the studies of A-Life have the ability not only to explain living systems but to create them.

These problems will be addressed from a philosophical standpoint by analyzing the assumptions made within the field of A-Life especially in its claims to the possibility of creating man-made life forms.

As the greater part of the work in A-Life is accomplished with the aid of computer hardware and software, it will also be necessary to determine if the computer is a suitable medium within which life could exist.

What is Artificial Life?

The scientific study of A-Life was officially named and initiated in September of 1987. On that date the first Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems was held in New Mexico at the Center for Nonlinear Studies, Los Alamos National Laboratory. Prior to this conference, the study of A-Life was just a vaguely formed concept floating around the minds and work of various researchers in the field of computer science (Levy, 1992). These researchers had become interested in the study of biological theories with the aim of applying biological laws and processes, such as evolution, to the production of self-replicating computer programs (Levy, 1992).

These scientists had all been working on separate projects and many did not even know of the existence of other researchers who might be interested in similar studies. This diverse group of unrelated thinkers met through the invitation of Dr. Christopher Langton, who had determined that many of the researchers' projects he had been interested in had the similarity of trying to apply biological processes to computing. So, in order to facilitate discussions and collaborative thinking on the subject he organized the above conference (Levy, 1992).

The purpose of this conference was to attempt to fuse the dispersed projects of unrelated researchers who nonetheless were pursuing similar ideas into a new field of scientific study which would help advance the use of computer models in the study of, and possibly the creation of, life forms in theoretical biology (Langton, 1989, p. xvi).

One of the fruits of this conference was the creation of a more detailed description of what the science of A-Life actually is. This definition is given in the preface to "Artificial Life, The Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems," which describes A-Life as follows:

> Artificial Life involves the *realization* of lifelike behavior on the part of man-made systems consisting of *populations* of semi-autonomous entities whose *local interactions* with one another are governed by a set of *simple rules.* Such systems contain *no* rules for the behavior of the population at the global level, and often complex, high-level dynamics and structures observed are *emergent* properties, which develop over time from out of all of the local interactions among low-level primitives by a process highly reminiscent of *embryological development,* in which *local hierarchies* of higher-order structures develop and *compete* with one another for support among the low-level entities. These emergent structures play a vital role in organizing the behavior of the lowest-level entities by establishing the context within which those entities invoke their local rules and, as a consequence, these structures may *evolve* in time. (Langton, 1989, p. xxii)

It must be noted that even though this definition has the feel of being the official manifesto of the science of A-Life, one must remember that this science is still only in its early development and that no hard and fast definitions or laws have come out of its activities as of yet. And even though Dr. Langton wrote the above definition of A-Life, he also states in the editor's introduction to the first issue of the journal Artificial Life that "artificial life is not yet ready to be constrained by quick and short definitions. . . . Artificial life is still in the process of defining itself, as is proper for any new discipline" (Langton, 1993, p.v). So we must show some caution in too harshly judging the initial offerings of definitions of A-Life. Still it is our duty as philosophers to critically analyze a new science as it is forming its self-identity, as this is precisely the moment when outside influences can beneficially affect a science.

After an initial phase, a science is likely to solidify around a core of ideas and forms of practice that are nearly impossible to change. This is because once a science identifies itself to a particular set of issues and problem-solving techniques, and once resources and grants are committed to specific studies, the science loses its initial openness and flexibility and at this point the science can begin to dogmatically cling to its "officially" stated ideals.[1]

With this in mind we can see that it is important for us to critique the early views of A-Life with the intent of helping it formulate its basic tenets and practices in order to foster an environment conducive to the fruitful study of A-Life. And as the recent feminist critiques of science have shown us, even if it is not possible to completely eliminate all forms of bias from science, it is still possible to work towards the goal of eliminating as much bias as possible in order to create less false and less distorted ideas about what a science studies.[2] These concerns must be addressed early in a new field of scientific endeavor so that mistakes made in other, similar sciences can be addressed before they are repeated.

---

[1]This idea is influenced by Langdon W. (1986). The Whale and the Reactor, Chicago: University of Chicago Press. And Paul Feyerabend's essay, How to Defend Society Against Science, in Philosophy of Science, (1988), E. D. Klemke, R. Hollinger, A. D. Kline (Eds.), Buffalo New York: Prometheus Books . Pg. 34.

[2]With this type of vigilance it may be possible to avoid some of the prejudices and biases that inherently creep into science from the society in which it exits. For a discussion of this idea see Harding, S. (1991). Whose Science? Whose Knowledge? Ithaca New York: Cornell University Press.

Dogmatism and the Sciences of the Artificial

In order to illustrate what I mean by the statements I have made above, I would like to take a brief look at the science of Artificial Intelligence. This science has been scathingly critiqued by some observers, the most prominent of whom are John Searle and Herbert Dreyfus. Whether or not one follows the arguments of these critics, it still remains that Artificial Intelligence is a science that is currently in a crisis.[3] The science of Artificial Intelligence has not been able to deliver on its promises of being able to completely model human intelligence by the turn of the century.

Perhaps the strong claims made by early AI researchers were made a little too prematurely. But, unfortunately, some rather strong claims were made early on and the science has not been able to deliver on its promise of completely modeling the human mind.

It is widely held that Alan M. Turing's famous paper, "Computing Machinery and Intelligence," seems to be the spark that ignited the imaginations of many researchers in the fledgling field of computer science as to the possibility of creating an artificially intelligent machine. Reading this paper one is led to believe that modeling human intelligence is not only theoretically possible but also practically achievable in only a matter of decades from the time of its writing in 1950 (Turing, 1992). Turing suggested that it would soon be possible to create a machine that could converse so well with another person that that person would have no idea that he or she was speaking to a machine. (One has to imagine that the two are conversing through a curtain or some other similar scheme which hides the machine.) Turing advanced these ideas beginning in the late thirties and, while there has been some success in modeling very specific

---

[3]For a discussion of this, see Freedman, D. H. (1994). Brainmakers. New York: Simon & Schuster.

instances of human intelligence, we have yet to see a fully autonomous machine capable of conversing with a human on a wide variety of subjects.

Critics such as Searle and Dreyfus blame this failure on what they see as fundamental flaws in the assumptions made by the science of Artificial Intelligence regarding what intelligence consist of, and what would be an adequate model of that intelligence. These arguments are generally critical of the reductionistic theories of simulation employed by Artificial Intelligence theorists. Most critics of Artificial Intelligence believe that no amount of time will allow for the completion of the goal of modeling human intelligence with a machine. Apologists suggest that researchers in Artificial Intelligence were just over-exuberant in making estimates about when it would complete its tasks, but that its eventual success (though maybe many years away still) is logically viable.[4] A third opinion, which has direct bearing on the subject of this paper, is the idea that the techniques used in traditional artificial intelligence programming may be flawed and that idea is in fact exactly opposite of what is necessary. This theory suggests that Artificial Intelligence has gone about the modeling of intelligence in the wrong direction. This theory is called Computationalism. Computationalism appeared on the AI scene in the 1970's in the work of David Marr (Freedman, 1994, p. 19).

Computationalisim suggests that rather then first modeling high-order intelligence such as logic and language we should start by analyzing lower-order abilities like perception, motor control, and reflexes "and then distill from this analysis a rigorous mathematical description of the processes that take place; in theory, such mathematical representations can then be implemented on a computer" (Freedman, 1994, p. 20).

---

[4]See Simon, H. (1994, July 1994). Interview, Herbert A. Simon. Omni, 72-89.

The computationalists argue that if we look at nature we see that evolution has spent millions of years working on solutions to some very basic problems of intelligent behavior whereas it has only been working on logic and language for a few hundred thousand years (Freedman, 1994). A well-known researcher in this field, Thomas Poggio suggests:

> AI [artificial intelligence] has been trying to construct a tower starting from the top; instead of digging right into reasoning and problem solving, perhaps Artificial Intelligence should be trying to construct the foundation on which such capabilities rest, as is apparently the case in living creatures. (Freedman, 1994, p. 19)

This view seems obvious enough, but it has taken years for this idea to take hold amongst the practitioners of Artificial Intelligence. Still this view is seen as a sort of maverick movement away from traditional Artificial Intelligence (Freedman, 1994, p. 22).

We can also see a similar situation arising in the study of Neural Networks. The theory behind Neural Networks was once widely thought of amongst the leading researchers in Artificial Intelligence as being inefficient and debunked. But the study of Nural Networks has recently made a powerful comeback and is now a serious challenge to the standard Artificial Intelligence techniques of pure symbolic logic programming methods.[5] In his book, *Turtles, Termites, and Traffic Jams,* Mitchel Resnick suggests that:

> The Early days of Artificial Intelligence, in the 1950s, were characterized by a diversity of approaches. Some researchers experimented with perceptrons and neural networks, networks of

---

[5]See Dreyfus, H. L. a. S. E. (1992). Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branch-Point. In M. A. Boden (Ed.), The Philosophy of Artificial Intelligence. Oxford: Oxford University Press. p. 334.

> simple computational elements. No single element was in charge of the network. Rather, solutions emerged based on interactions among the distributed elements. (Resnick, 1994, p. 16)

Resnick contends that in the 1960's the initial enthusiasm for decentralized systems was replaced with a paradigm that attempted to model the mind with strictly centralized architectures appearing in planning systems like Strips and in Expert Systems (Resnick, 1994, p. 16). While it would be wrong to say that nothing of worth has come out of the study of traditional AI, it would seem that the hard shift from distributed architectures to a more or less dogmatic pursuit of only centralized architectures may have resulted in the science of Artificial Intelligence falling short of its stated goals and time lines.

While the centralized architecture is beneficial for creating expert systems and chess programs, etc., it has proven much too brittle in its ability to model the basic skills such as movement, vision, and coordination, which living creatures accomplish with ease, and in which distributed architectures seem to make more progress.[6] In fact Artificial Intelligence guru Marvin Minsky has acquiesced to the growing interest in distributed architectures and has even begun to shift to the decentralized camp as evidenced by his Society of Mind theories (Freedman, 1994, p. 22). The dogmatism of the early government-funded Artificial Intelligence research was a near fatal flaw in the science of Artificial Intelligence.[7]

It would seem that in the early stages of the development of the science of Artificial Intelligence many ideas were hastily abandoned only to be exhumed decades later; this is exactly the type of situation that can be avoided by a careful philosophical examination of a new science right from the beginning. Of course poorly done or

---

[6]As evidenced by the robotics work of Poggio and the computer modeling of Karl Sims.

[7]See Freedman, D. H. (1994), for a thorough discussion of the history of this debate in AI.

dogmatic philosophical investigations of a new science can impede the progress of that science quite seriously. One only has to remember the trials and tribulations of the early scientists like Galileo, who were terribly impeded by a dogmatic approach to Aristotelian doctrine as it was understood at the time. This is clearly undesirable and we as philosophers have to approach our investigations of new sciences with careful and precise, yet open minds. In the case of Artificial Intelligence a philosophical study of the theories that drive the quest for artificial intelligence is still in order, and if it had been thoroughly pursued earlier it might have benefited Artificial Intelligence greatly.

One must notice how the science of Artificial Intelligence was formed out of the logic theories of Lebnitz and the philosophies of the last century that posited a world that is entirely knowable and harnessable. These concepts are debatable and any undue reliance on them in the field of Artificial Intelligence needs to be examined and corrected. But that is a task we will for the most part put aside. This paper will concern itself instead with some of the philosophical underpinnings of A-Life with the intent of helping it bypass the philosophical quagmire that Artificial Intelligence has sunk into so deeply. In this way we can do our part to help advance an important new science along the best foreseeable path towards a successful future.

CHAPTER 2

What Can Be Gained from the Study of Artificial Life?

A-Life is not merely some curiosity that has no bearing on the real problems of the world. A-Life is a significant topic that has the potential to affect many areas of study including, art, science, and philosophy. The first and most obvious benefactor of the study of A-Life is the field of theoretical biology.

Claus Emmeche, a theoretical biologist, has already begun to take notice of the breakthroughs and challenges that A-Life offers to his field of study. While Emmeche is somewhat guarded in his belief in the strong claims made for A-Life, he still feels that "theoretical biology could benefit from new approaches to its subject matter in order to make progress as to the general aspects of living systems" (Emmeche, 1992, p. 467). He views A-Life as a "postmodern science," in which the older science of biology is deconstructed and viewed again, thus revitalizing the science of biology (Emmeche, 1994, p. 161). This is a very intriguing claim and we will be looking at it closely near the end of this paper.

A-Life also promises to deliver great rewards in the field of computer programming. This new style of programming uses the "genetic programming paradigm" to attempt to harness the forces of natural selection for use in creating code for programming (Koza, 1991, p. 603). The concept for the genetic algorithm was conceived by John Holland in the late 1950's at the University of Michigan in the Logic of Computers Group (Levy, 1992, p. 158). After about fifteen years of toying with the concept and teaching it to his graduate students, the idea was finally applied to a real problem by one of Holand's graduate students, David Goldberg, who successfully used the process to solve problems in allocating resources in natural gas (Levy, 1992, p. 161). The genetic algorithm is so named because it

> genetically breeds populations of computer programs to solve problems. In the genetic programming paradigm, the individuals in the population are hierarchical compositions of functions and arguments of various sizes and shapes. Increasingly fit hierarchies are then evolved in response to the problem environment using the genetic operations of fitness proportionate reproduction (Darwinian survival and reproduction of the fittest) and crossover (sexual recombination). In the genetic programming paradigm, the size and shape of the hierarchical solution to the problem is not specified in advance. Instead, the size and shape of the hierarchy, as well as the contents of the hierarchy, evolves in response to the Darwinian selective pressure exerted by the problem environment. (Emmeche, 1994, p. 161)

This style of programming promises to become a very important breakthrough in the field of computer science. John R. Koza, the Stanford computer scientist who wrote the above quoted passage, believes that "entire computer programs can be genetically bred to solve problems in a variety of different areas of artificial intelligence, machine learning, and symbolic processing" (Emmeche, 1994, p. 607). Koza lists a number of programs which he has created using genetic algorithms which have solved some minor but tricky problems in the fields of robotics, mathematics, and artificial intelligence.

What is most interesting about these programs is that they are not created by a programmer in the traditional sense. In traditional programming the structure of the program is imposed by the programmer. This means that the programmer has to conceive of every variable and nuance to the proposed problem. This is often a insurmountable task. In the genetic algorithm paradigm the structure of the program emerges mostly on its own with the persuasion of the programmer. In other words, instead of a programmer attempting to plan out every aspect of a program before hand and then writing the code, in the genetic algorithm paradigm the programmer applies thousands of small almost random programs to the task of solving the problem, allowing the most successful to combine or "breed" and then applies the newly created "offspring" to the problem again; this process is repeated over and over until the program outputs a correct solution. When a program is written with genetic algorithms many surprising things happen. To the layman the results of using the genetic algorithm paradigm seem similar to putting all the parts of an automobile into a box, jumbling it around and then having the parts happen to all come together to form the finished product. Logically it shouldn't work. But this analogy is lacking. The process wouldn't work in the case of the automobile because the various parts used to construct it are not interchangeable or general enough. One couldn't rub some nuts and bolts and sheet metal together and expect a transmission to result from the process. But in the informational environment of the computer the "parts" are extremely interchangeable and general, and the process used to combine them is much less chaotic and random then "shaking" a box full of parts. Even a rudimentary knowledge of computer science reveals that at their deepest levels all computer programs operate using simple on and off switches logically arrayed; one can see that these "parts" or logic circuits are completely interchangeable, a fact which has caused the digital computer to be considered a universal machine.[8]

We can see that it is not impossible then to create a program that will properly mix various functions in a program yet still facilitate the proper aligning of these switches through the process described above. Creating generation after generation of mostly ineffectual computer programs in order to create an optimum solution to the

---

[8]See, Rucker, R. (1987) Mind Tools, Boston: Houghton Miffin and Company.

problem at hand is a large and tedious problem to be sure, but the computing power available today is sufficient to crunch through the generations at a significant speed, which begins to make genetic algorithms a profitable alternative to conventional programming.

The genetic algorithm is one of the fundamental building blocks of A-Life. Clearly John Holland and John Koza have shown that the genetic algorithm has the possibility to provide immense practical benefits for computer programming. But what is interesting philosophically is to determine whether the process is *metaphorically* similar to Darwinian evolution as it occurs in nature or whether it *is* another example of Darwinian evolution in nature. Or put another way, Is the environment of digital information processing an interesting way to model nature, or is it part of the fundamental structure of nature? We will take up this discussion in the last chapters of this paper.

A-Life also impacts the arts and humanities. One of the immediate benefactors in the field of A-Life programming is that of computer art. Cellular automata are already in use in computer animation where they are used to generate moving patterns and textures. As cellular automata are one of the building blocks of A-Life, it is only natural that full-blown A-Life programs will be applied to computer art. The geneticist Richard Dawkins is an outspoken advocate of A-Life.[9] One of his early experiments using A-Life involved a simple program by which he hoped to show how evolution can achieve the complexity we see in contemporary life forms (Levy, 1992, p. 172). His program caused the computer to randomly generate shapes on the screen beginning with a single pixel. It would then randomly mutate the "genes" representing this shape and show a few of the results to the user. The user then would choose the most pleasing

---

[9]See, Dawkins, R. (1995,). Revolutionary Evolutionist. Wired, 120-186.

"child" and the process would continue. Dawkins only expected to be able to create vaguely tree-like shapes but to his amazement he was able to create vastly more complex forms (Levy, 1992, p. 173). Many of his creations, which he calls "biomorphs," look a lot like insects and they are indeed very complex shapes. This technique has since been used by artists with more sophisticated computers to create more interesting works of art.

The best known artists to use these techniques are perhaps the team of Stephen Todd and William Latham. They have been using A-Life programming and powerful computers under a grant from IBM to evolve complex and striking computer graphics using a style of art they call "Evolutionism. "Evolutionism divides the artistic process into two parts: in the first part, the artist creates an artificial world by defining systems and structures for form and animation generation; in the second part, he works as a gardener within this world, using aesthetic judgment to breed artworks" (Latham, 1992, p. 207). Latham's works depict wildly imaginative forms which one could imagine discovering in some alien ecosystem or perhaps seen in a microscope pointed at a sample of pond water.

These images are created by evolving and combining simple shapes and with the help of a computer vastly detailed and naturalistic designs can be created. Latham uses the rules of biology as an inspiration to his work; "the natural basis of the rules gives the works an organic realism, and the artist's imagination imposes a surrealist feel" (Latham, and Tod, 1992, p. 504). One can see that this is a visual application of the genetic programming techniques described above, and as both are only in there beginning stages we have yet to see the limit of what they might develop.

This type of art work is controversial to be sure because the artist does not seem to be in control of the work. Of course an artist is never entirely in control of his or her medium, but artists like Latham seem to be interested in letting as much of the structure of their work emerge from the processes of the computer as possible. In fact, there is some talk about using a similar type of system to allow the viewer of the work to interact with it and change the piece as it is viewed (Bonabeau, and Theraulaz, 1994, p. 322). Even if it is not art as we know it, it is still an interesting phenomena and it will no doubt prove fruitful to computer animation.

A-Life also stands to affect the field of philosophy. The philosopher and cognitive scientist Daniel Dennett has suggested that there are two different ways in which philosophers can deal with the emerging science of A-Life. The first way is to see it as a new way of doing philosophy, and the second is to see it as a new phenomenon by which to analyze using the traditional philosophical methods (Dennett, 1994). Dennett's bias is towards the former possibility. He sees A-Life as a powerful new tool for philosophers to use. Dennett holds that:

> Philosophers have always trafficked in thought experiments, putatively conclusive arguments about what is possible, necessary, and impossible under various assumptions. The cases that philosophers have been able to make using these methods are notoriously inconclusive. What "stands to reason" or is "obvious" in various complex scenarios is quite often more an artifact of the bias and limitations of the philosopher's imagination then the dictate of genuine logical insight. Artificial Life like its parent (aunt?) discipline, Artificial Intelligence, can be conceived as a *sort* of philosophy--the creation *and testing* of elaborate thought experiments, kept honest by requirements that could never be imposed on the naked mind of a human thinker acting alone. In short, Artificial Life research is the creation of prosthetically controlled thought experiments of indefinite complexity. This is a great way of confirming or disconfirming many of the intuitions or hunches that otherwise have to pass as data for the sorts of conceptual investigations that define the subject matter of philosophy. Philosophers who see this opportunity will want to leap into the field, at whatever level of

> abstraction suits their interests, and gird their conceptual loins with the simulational virtuosity of computers. (p. 292)

Certainly there is much to take issue with in the above passage. For instance the implications that all philosophers use dubious thought experiments and that Philosophy could stand a more thorough realigning with scientific methods, are both old criticisms which can be argued against. Questions like the above will have to be put aside for now, but some of Dennett's claims do bear directly on the issue at hand.

In this paper we are trying to develop a clear picture of what the phenomenon of A-Life is. The answer to that question will determine the types of projects for which A-Life would be a good tool to use. For instance, if Emmeche's claims are correct, and A-Life is a type of simulacrum then that fact will color quite seriously any results obtained through philosophical experiments using A-Life as a tool for testing hypotheses. So philosophers should not uncritically adopt A-Life as a tool without a thorough understanding of its inherent biases.

Dennett lists a number of interesting problems to which A-Life testing can be applied such as "traditional philosophical issues in the philosophy of biology, of science, of mind, and even metaphysics and ethics . . . [for example,] are Hobbesian just so stories about the possibility of the evolution of cooperation defensible" (Dennett, 1994, p. 292)? All of the subtopics of philosophy he mentions are interesting and I agree that A-Life may be a good tool to use in studying them but I feel that some caution must be exercised in the use of the tool of A-Life in the study of philosophy for the reason given above.

Another philosopher who is doing pioneering work in the philosophy of A-Life is Mark A. Bedeau. In his article "*Philosophical Aspects of Artificial Life,*" he outlines fourteen philosophical questions that he believes experimenting with A-Life programs can help us answer (Bedeau, 1992, p. 494). These questions deal with defining life, clarifying the concepts behind functionalism and emergence theories, and the ethics of creating living things (Bedeau, 1992, p. 494). Bedeau also believes that the fields of A-Life will eventually affect the study of philosophy in as profound a way as Artificial Intelligence. I agree with him, as A-Life can be seen as approaching the same sorts of issues as Artificial Intelligence, the difference being in the method of solving the problems. Artificial Intelligence begins with attempting to simulate extremely complex systems and moving down, and A-Life begins with simulating simple systems and moving up the ladder of complexity.[10] Bedeau also defends the involvement of philosophy in A-Life as follows:

> The philosophy of artificial life is not merely a derivative, second-order gloss on first-order A-Life science. In fact, most of the questions . . . [asked by philosophers] are of direct and fundamental concern in artificial life *science*.. However unclear the relationship between A-Life science and A-Life philosophy might be, the two are surely closely related, and in time will co-evolve. (1992, p. 494)

So we can see that the study of A-Life is going to be extremely fruitful. The main problem though is going to be how to correctly apply to the world we live in what we find from studying A-Life. Later on, I will try to show that all of what is done in A-Life is a simulation, and that being the case, we have to be very clear as to the value and proper use of simulations in order to correctly judge the value of findings in A-Life studies.

It is appropriate now to take a somewhat detailed look at where the science of A-Life comes from and also to try to become familiar with how A-Life works. The best way to do this is to look at the work of John von Neumann.

---

[10]Taken from a conversation with Rudy Rucker.

CHAPTER 3

The Artificial Life of John von Neumann

In order to understand what A-Life is today we need to look at what ideas and theories A-Life has evolved out of. The most important of these seminal ideas is undoubtedly the theories of John von Neumann. John von Neumann was a well-known mathematician and scientist. He was part of the pantheon of the great thinkers such as Einstein, Gödel, Schrödinger, Turing, etc., in mathematics and physics of the early part of this century John von Neumann is remembered principally for his contributions to the development of the atomic bomb and for his work in the development of the digital computer. In fact, his contributions to the computer were so important that computer processors are commonly referred to as von Neumann processors (Levy, 1992, p. 13). Among the many projects for which he is less well known is his interest and work in the theory of self-replicating machines. It seems that his motivation for these studies was that he wanted to create a theory which would prove that a mechanistic explanation for life was not fundamentally impossible (Sigmund, 1993, p. 16). John von Neumann was

> infatuated with the similarities of the computer--or, more precisely, with what this machine could become--and with the workings of nature. His goal was to create a theory that would encompass both biologies, natural and artificial. (Levy, 1992, p. 14)

Self-replicating machines, or automata, are an essential first step towards proving a mechanist theory of life. Unlike some of the more recent A-Life researchers, von Neumann seems to have considered self-reproduction as the prime ingredient for determining whether or not a system could be considered to be alive (Emmeche, 1994, p. 51). Von Neumann felt that "creating offspring is a prime common-sense criterion for determining whether something is alive" (Emmeche, 1994, p. 16). The existence of such machines would not clinch the issue but without them the theory is literally and figuratively dead.

It is appropriate here to look briefly at the specifics of von Neumann's theory since "von Neumann's approach has been, ever since the early 1960s, the dominant if not the only approach to artificial life" (Sigmund, 1993, p. 23). Von Neumann's theory of self-replicating automata is purely formal, based on logic and mathematics. He never intended on building the machine he describes; he only wanted to prove that it was not unthinkable to believe such a machine might be possible (Sigmund, 1993, p. 17, and Emmeche, 1994, p. 51).

Mechanical Sex

Von Neumann knew that not just any form of reproduction would do. For instance one might say that a factory that produces lawnmowers is, strictly speaking, a machine that produces other machines, yet this is not an instance of reproduction because the machines produced are much less complex than the machine doing the production. In order for a system to be able to reproduce it must be able to form a copy of itself that is at least as complex as the original and that can go on to create more copies of itself (Sigmund, 1993, p. 17). Von Neumann also distinguishes between trivial and robust self-replication. Karl Sigmund gives the example of a robot wandering aimlessly around a warehouse full of other dormant robots which can be turned on by the original robot any time it bumps into one. One can imagine that after a time the warehouse would be teaming with active robots as each newly "created" robot bumps into more dormant robots and turns them on, etc. (Sigmund, 1993, p. 161). But this is not very much like what happens in nature and could not be considered an example of life. Von Neumann calls this trivial self-reproduction and likens it to the growth of a crystal in a solution:

> The crystal grows because the right atoms join up; if it breaks, each part grows on its own. This is self-reproduction of a sort, but hardly deserves to be labeled *life*. (Levy, 1992, p. 161)

This trivial form of self-reproduction is possible even in simple machines. The mathematician L. S. Penrose proved this empirically in the mid-1950's. Penrose and his son Rodger built simple machines out of shapes cut out of plywood which could attach to one another in various ways. When these shapes were placed in a vibrating box along with a *seed* of pre-configured pieces, the parts would bump together and after a time form many replicas of the seed configuration. Penrose was inspired to do these experiments after hearing the theories of von Neumann, but he was never able to move beyond this simple form of self-reproduction (Penrose, 1992). It is interesting to note that after Penrose, no one has pursued the study of mechanically produced self-organization until recently. Researchers in the Department of Mechano-Informatics at the University of Tokyo have taken an interest in Penrose's studies and are conducting experiments using blocks fitted with magnets randomly mixed in a moving container in order to attempt to find a mathematical formula[11] to describe the tendency for systems to self-organize.[12]

---

[11]This mathematical formula consists of the adoption as state variables of the quantities of each intermediate product of the simple mechanical shapes involved in the experiments. The systems behavior is reduced to a set of difference equations with a small degree of freedom, in a fashion similar to that used in chemical kinetics or in population dynamics, however its use in self-assembling systems is unique. This

Still, Von Neumann wanted something more than this. He wanted his automata to be able to create their offspring out of basic raw materials and he wanted the automata to be able to construct their offspring much as a new creature gestates in a womb or egg. He wanted to model true reproduction not just spontaneous organization, and to von Neumann's mind the key to giving this ability to machines lay in their ability to process information.

In order to reproduce in the robust manner in which von Neumann wanted, the machine he felt would have to have some sort of plan or blueprint contained in itself that it could use as the design for the new construct. One has to remember that when von Neumann was conceiving this idea no one had discovered how an actual living biological cell reproduces itself, and therefore the concept of DNA and coded genetic information was not an accepted scientific fact. So in some sense von Neumann was a visionary when he suggested that the fundamental ingredient for life is the ability to pass on information between generations (Levy, 1992). The necessity of the machine to be able to manipulate information this way led von Neumann to place a computer as the central component to his automata. Von Neumann of course was no stranger to computers or what they could theoretically do.

Von Neumann began by describing a thought experiment where one is to imagine an automaton floating in a sea or lake of some sort which happens to be stocked with simple basic parts or "organs" as von Neumann calls them (Emmeche, 1994, p. 54). The automaton floats around and encounters these basic parts and attempts to construct a copy of itself. One could criticize this as an example of trivial self-reproduction, but von Neumann would be able to counter by saying that if one

---

experiment is described in Artificial Life, Vol. 1 Number 4, Cambridge Massachusetts: MIT Press Journals summer 1995. Pg. 413-427.

[12]Also in Artificial Life, Vol. 1 Number 4, Cambridge Massachusetts: MIT Press Journals summer 1995. p. 413-427.

looks at actual biological life, the splitting of a cell for instance, one sees that it also needs a highly specialized environment in which to thrive. The living cell must have access to complex molecules like fats, sugars, proteins, and specialized sources of energy, without which the cell will die (Sigmund, 1993, p. 17). So we can see that von Neumann's thought experiment is fine up to this point. Claus Emmeche gives a good brief description of the workings of this automaton in his book, *The Garden in the Machine.* Emmeche's description is useful to look at here (p. 54-55). He describes the machine as consisting of four parts: A, B, C, and D:

> A is a general constructor-automaton or "factory." It produces an output X when given an instruction or "logical description" b(X) of the desired output. From this description, the factory chooses the suitable components in the pond of raw materials in which it moves around. Of course, the description must first be transmitted to A.
>
> B is a general copying automaton or duplicator, which with a given description b as input delivers the description b, plus a copy called b', as output.
>
> C is a controller, which delivers to B an instruction, b(X), for double copying. C then appends the first copy to A in order to execute the construction of the described "product" X (here von Neumann conceived the construction as occurring during the destruction of the copy of the instruction). Finally, the controller attaches the remaining copy b(X) to A's output X, and then releases this (X + b(X)) from the machine A+B+C, retaining the original itself.
>
> D is an instruction of a particular kind. D is a description that enables A to produce exactly A+B+C. In other words, D is the machine's self-description, D= b(A+B+C).

One can see that this logical description of self-replication can be easily applied to the division of living cells (Emmeche, 1994, p. 56, and Sigmund, 1993, p. 19). It is even possible to allow for evolution via mutations.

If we postulate that automata of this type could have extra components such that the automaton might be described as (A+B+C+D+E) then a mutation in the description (D) such as "change E to F" for instance would not cause the automaton to be infertile, as none of the essential duplicating functions (A, B, or C) have been altered but the new offspring, (A+B+C+D+F), is a new type of automaton which may be better then its parent (A+B+C+D+E) (Emmeche, 1994, p.56). The reproduction described above is strictly asexual but with one minor change von Neumann's automata can enter the wild world of sexual reproduction. This sexual reproduction could occur if two automata met and some how swapped parts of their individual descriptions D, with each other creating a hybrid description D, which they could then reproduce using the systems A, B, and C.

So there we have it, a straightforward and logically compelling description of the possibility of self-replication automata. Von Neumann died before he could carry his theories further but others came along and picked up where he left off.

<u>Where No Automaton Has Gone Before: Virtual Colonialism</u>

Of course von Neumann never really intended to physically build his automata but others have not been so cautious. For instance, a ten-week program sponsored by NASA in 1980 in order to determine the role of advanced automation and robotics devices in future space missions produced a plan on how von Neumann automata might be used to colonize the moon and eventually the universe (these men certainly weren't afraid to think big) (Levy, 1992, p. 34). And prior to that, Edward F. Moore proposed that we build "Artificially Living Plants" which would be complex von Neumann automata that would comb the seas of earth like mechanical whales harvesting the abundant raw materials found in the sea and creating replicas of themselves. The practical reason for doing so would be that one could program the machine to produce a little more of some valuable material than was necessary for self-reproduction, e.g., fresh water, magnesium, gold, etc., thus yielding untold profit for the initial investment of a single self-replicating factory (Levy, 1992, p. 32). One could easily imagine the environmental disaster that such machines would create and luckily no one has taken the idea seriously yet.

On the other hand, the ideas generated at NASA, although officially ignored now, may be acted upon some day. All it would take is the initial investment of a single probe launched into space capable of self-replication. This probe would use the floating bits of matter and solar energy which abound in the universe to slowly build a working replica of itself. The probe and its replica would then go on to build another generation of replicas. This would continue to go on indefinitely and with some luck, and a lot of time, there could be a tremendous amount of these things floating around the universe. The idea is that these probes would be similar to others that we have sent out (e.g., Voyager), and that they would send back to us messages about all of the information they gather about the universe. Unfortunately, even if one could get this project to work it would be a very immature thing to do. There is no way to know what kinds of repercussions these probes might cause. It's like a bad science fiction plot; imagine something like this launched by some alien race millennia ago landing here and beginning to "process" the "raw materials" of your house. These things would be like a universal virus.

This potential horror story did not go unnoticed by the think tank that conceived the idea. In fact the originators of this idea speculated that it might not be a good idea to ever develop this technology as the mutated progeny of the original probe might some day evolve intelligence and compete with us directly for dominance of the universe (Levy, 1992, p. 42). Others present at the conference suggested that it is somehow our duty to develop this technology as these intelligent star probes would be essentially the next evolutionary step of mankind's manifest destiny to the stars.[13] One wonders how much NASA spent on this project. Still it is not too far fetched to believe that some government might take an idea like this seriously some day. In 1983 the rumor among NASA scientists was that Ronald Reagan was likely to sign off on a program to start researching self-replicating lunar factories. Instead, he initiated a much more loony project, the infamous "Star Wars" proposal (Sigmund, 1993, p. 23).

A Purely Formal Model

Putting aside the impracticability of implementing von Neumann's automata we are still left with an interesting description of the mechanizability of self-reproduction. But Von Neumann himself was dissatisfied with his thought experiment. It was not abstract enough for his taste. He wanted a purely formal argument. He knew that when presented with his model, "engineers would have no complaints, but a philosopher might still harbor the suspicion that somewhere, concealed . . . some unknown vital force or supernatural agency was at work" (Sigmund, 1993, p. 22). Von Neumann needed to remove his thought experiment from the problematic world of actual robotics and into a more formal and mathematical realm.

He found that realm in the work of his friend Stanislas Ulam. Ulam had been working on using the computer as a tool of discovery in the field of mathematics and had developed cellular automata. Cellular automata are best described by example. Imagine a grid and each cell in that grid has a potential value. This value may be expressed numerically and can be as simple as 0 or 1 or as complex as you would like.

[13]This oddly utopian idea that machine life will be able to succeed where human frailty has not will be a constant companion of most of the talk about A-Life either consciously or subconsciously.

To determine the value of any specific "cell" one applies a specific preset formula which must fit with the following description. The rules must be:[14]

1. Parallel - the individual cell values are updated independently of one another, and all at the same time. (Note that a computer only simulates this, as the von Neumann architecture used on most computers is serial in nature, but a high speed computer can simulate parallelism fairly well by using buffers.)

2. Local - when the cell is updated, its new value is based entirely on its old value and the values of its nearest neighbors.

3. Homogeneous - every cell is updated according to the same rules. Usually the value of the cell and its nearest neighbors are combined in a logical or algebraic formula, or they are used to locate an entry in a preset look-up table.

The most common way one encounters cellular automata is by display on a computer monitor. In this case the values for the various cells are represented by different colors, and as the program for the cellular automata runs, the screen changes in complex and interesting ways. There are many different types of cellular automata and they can be used to model "physical, biological, and sociological phenomena" (Rucker, 1989, p. 2). What interested von Neumann the most about this new tool was that one could use the cellular automata to create small environments that were perfectly known. That is to say that all of the rules and laws of these environments are under the complete control of their creator. Von Neumann hoped that this would finally be the answer to creating a strictly formal self-replicating automaton (Sigmund, 1993, p. 22). Von Neumann developed a cellular automaton in which each cell could have up to twenty-nine states (Sigmund, 1993, p. 23). He was preparing a book to

[14]The following is paraphrased from Rucker, R. (1989). Cellular Automata Laboratory (Version 1). Sausalito, California: Autodesk. p. 2.

outline how this automaton could be seen as an example of a self-replicating machine, but he died before his book was finished.

This book was finished by his student Arthur Burks, who plugged some of the holes in the theory and proved that it could in principle be run on a computer (Emmeche, 1994, p. 58). At this point the field of artificial life studies began its long love affair with computer simulations.

We should take a closer look at von Neumann's automaton. It consisted of

> a two-dimensional cellular automaton, who's squares could assume some twenty-nine different states, one of them being earmarked as the 'empty' state. He laid down the recursion rules governing the transitions from one generation to the next, thereby specifying all that could possibly happen. And he devised a pattern which reproduced itself. It consisted of some two hundred thousand cells, and was again composed of two parts: (1) a construction unit including a universal computer, and (2) an instruction programming the replication. (Emmeche, 1994, p. 58)

When the automaton is run on the computer it grows an "arm" into an empty region of the two-dimensional grid. There it grows a replica of itself still attached via the arm. Once a replica of itself has been built it transfers the replication program to the new automaton and the original automaton withdraws the arm.[15]

Unfortunately this automata is very complex and since its inception many others (Arthur Burks, E. F. Codd, Christopher Langton) have devised simpler automata based on the same general ideas. These new models are specifically designed to run on a computer and nearly all A-Life simulations follow this general trend.

---

[15]This description is paraphrased from Sigmund, K. (1993). p. 22.

CHAPTER 4

The Birth of Artificial Life

We have seen how some of the basic background ideas for A-Life were gestating in the work of thinkers like von Neumann and Burks. Of course these mathematicians were not the only ones responsible for laying the ground work for A-Life; for instance, the mathematician John Conway was another one of the prominent early figures in the development of A-Life and later we will look at some of his work on the subject. But right now it is important to look at a specific example of how von Neumann's ideas were used to launch the new study of A-Life. As I have mentioned before, Christopher Langton has been quite instrumental in formulating the basic research questions that underlie and shape the study of A-Life. His first foray into studying A-Life (although the field had no name at the time) was programmimg a cellular automaton based on the ideas of von Neumann.

It is an experience both exciting and oddly disturbing to watch Langton's *"Loops"* (as his automaton is popularly known) automaton running on a good computer with a color screen. The simulation begins with a blank screen in which a single square automaton is placed in the center. Instantly the automaton begins to extend an "arm" into the blank space around it. This arm "grows" until it has formed an exact copy of the first automaton; then it detaches itself from the "parent. At this point both automata start this process again until each is surrounded by offspring. Once all of the space around the automaton is filled with offspring then the automaton "dies. After a few minutes the screen is filled with these automata and when there is no more room the simulation ends. In our age of relatively sophisticated computer animation these automata might at first glance appear crude but when one realizes that what is happening on the screen is not a prerecorded movie but rather the graphical representation of very simple rules reiterated over and over, then we can begin to see the attraction that fascinates those who study A-Life.

> These embedded self-reproducing loops are the result of the recursive application of a rule to a seed structure. In this case, the primary rule that is being recursively applied constitutes the "physics" of the universe. The initial state of the loop itself constitutes a little "computer" under the recursively applied physics of the universe: a computer whose program causes it to construct a copy of itself. (Langton, 1988, p. 29)

With a little imagination one can see that what is happening on the screen looks like primitive cellular growth just like one might find in the growth of a colony creature like corral. But of course just because something looks like it is alive doesn't mean it is alive. Still Langton has succeeded in creating a working representation of von Neumann's theories.

Of course Langton has seriously altered the initial twenty-nine state automaton of von Neumann. Langton's *"Loops"* automaton has only seven states or values that each cell can posses; Langton also dropped von Neumann's belief that the automaton should contain the ability to be a universal computer. This greatly simplifies the model. Langton's intent in creating this automaton was to create a sample of self-replication. His reasoning is that "it is highly unlikely that the earliest self-reproducing molecules, from which all living organisms are supposed to have been derived, were capable of universal construction, and we would not want to eliminate those from the class of truly self-reproducing configurations."[16] This seems to be a

---

[16]Langton as quoted in (Levy, 1992, p. 98).

reasonable claim and as long as one is able to accept the claim that natural physics and computational manipulation of information are synonymous, then we can see that Langton has made the first steps towards creating living, self-replicating machines. But as philosophers we certainly can't let an assumption like that go by unchallenged. This is what I intended to do in the following sections.

Real Qestions about Artificial Life

In our discussion of A-Life so far we have come across a few questions which deserve to be carefully considered from a philosophical viewpoint. The first of these questions has to do with determining whether or not the computer is an environment in which life can be sustained. To answer this question we have to come to some conclusions as to what a reasonable definition for life would be. Once we have this we have to then look at exactly what type of environment the computer is: its structure, its essential qualities, and its complexity. With this in hand we can then compare that environment with natural environments in which life clearly flourishes. Through this comparison we can then determine if the mere flow of information as embodied in the digital computer is a suitable environment for the existence of life forms.

The second question our discussions above has brought out concerns the concept of simulation. It is clear that A-Life is at the very least an attempt at simulating complex living systems. The question then is whether an extremely accurate simulation is significantly different then the thing simulated. The answer to this question would seem easy but it has proved to be still troubling to some. This point will be addressed but the results of this discussion are merely an introduction to a more interesting question which is whether or not A-Life is only a simulation of truly living things or whether it is a step towards some other kind of interesting entity. Put another way: if

A-Life is not life as we know it, could it not still be an entirely new type of being, or life as we do not know it? And even if this wild claim falls to our well-reasoned scrutiny we still have to look at Emmeche's intriguing claims that A-Life is a post-modern science whose primary value is that it challenges and deconstructs the traditional views of biology and puts them to a useful and much needed test. The following sections will attempt to answer these questions and in so doing we will put forth some of the main claims of A-Life as they are currently presented by current leaders in the field, such as Christopher Langton and others. This discussion will hopefully add something useful to the effort to build a solid philosophical underpinning to this new science as we discussed in the introduction.

CHAPTER 5

Towards a Theory of Life

I believe that this may be an extremely difficult theory to construct. And as we will see it may not be possible to completely form a theory of life using traditional scientific techniques. It seems that when we apply our traditional mental tools to the question, "What is life?" we are left with disappointing and inaccurate conclusions. This may be due to the fact that we traditionally use the tool of analytic thinking when we approach such a problem, and that that tactic may not be entirely useful when approaching this particular problem. The first step when using an analytic thought process is to isolate the thing we want to study from the "noise" of its surroundings, then we reduce the problem down to its smallest parts, and then once we have understood the essential basics of the problem at hand we then conclude that we know its true essence. Or so the story goes.

The analytic process has been famously rewarding in fields like physics and inorganic chemistry but has not produced results in life sciences to the same level as those in the inorganic sciences. This is not to say that there has been no significant discoveries in those fields using this method, I am only claiming that those discoveries have not been as broad reaching and all inclusive as they have been in, for instance, physics; we are still waiting for our "Newton of the grass blades. It will be a long wait. It may be that the reductionist methods of traditional science are just not the right tools to use on the problem of discovering what life is.[17]

[17]The following ideas where influenced by a talk given by Christopher Langton on September 25 1996 which took place at the Hewlit Packard offices in Palo Alto California, as well as personal discussions with him after the event.

Traditionally, when we want to know what a frog is we can catch one and dissect it. This allows us to make some determinations about what the internal structure of the frog is and we can pick apart the creature and make some reasonable claims about the biomechanics of how it gets the ability to jump, for instance. Still we have really only touched the surface of what a frog truly is. The frog on our table is one member of a certain species of a specific genus which, in turn, is related to further broad categories in a complex web of genes. We are forced by this system to generalize from a small sample all of the important qualities of the larger population of frogs.

This, of course, is well known and well adapted for problems of the scientific method. This problem has been known since the earliest days of science but even with this handicap, science has been able to give us valuable insights into the world. Given certain initial conditions, we can predict the orbits of the planets with great accuracy but when it comes to predicting the behavior of any living creature we have significantly less luck. Living things seem to respond chaotically and they lack the monotonous regularity observed in more mechanical systems.

The basic problem of studying specific life forms like our frog is compounded when we realize that frogs in general are members of an extremely complex web of relations which make up the environment of the pond, for instance. The pond in turn is one of many specific niches that compose the larger environment of a forest or marsh or waterway which is a system consisting of living and non-living things, such as other animals and the weather, etc., and is also nestled within a larger classification of environment. In just a few short, logical steps we have gone from studying a specific frog to having to study most of the known world. This combinatorial explosion of necessary data seems to me to be totally insurmountable using the traditional reductionist method. Or, as the biologist John Stewart puts it:

> It is certainly possible for an observer to distinguish between an organism and its environment. Nevertheless, in this case, the "environment" becomes an "ecological niche, and it is not possible for the observer to specify what the niche "is" without reference to the organism inhabiting it -- and reciprocally, the organism cannot be fully specified without reference to the "niche. Thus, in this sense also, object and subject are intrinsically inseparable. It follows that this approach is indeed non-objectivist. (1992, p. 478)

So we can come to know some specific facts through the reductionist method, such as how may digits a certain species of frog has on average, but we are limited, for example, in our knowledge of how that species of frog interacts with the environment and what its extinction might mean to the other creatures it shares its environment with. The traditional reductionist method for studying the basic question of what life is seems to be less than conclusive.

If we go back to the frog example we can see that soon after the dissection begins the frog is no longer a living thing; through this method we can learn allot about the minimum conditions for sustaining life but we get no closer to defining what life is. Life is a property of very complex organisms and even at its most basic level, the level of bacteria and viruses, we are still dealing with a problem many orders of magnitude more complex than a falling ball. In order to study living phenomena we have to study them using techniques that recognize this complexity and deal with the fact that we have to study a living system as a whole and that any attempt at reducing the system changes it entirely. This is the main thrust behind the sciences of complexity of which A-Life is a part.[18] These sciences are synthetic in their approach

---

[18]See, Lewin, R. (1992). Complexity: Life at the Edge of Chaos. New York: Macmillan Publishing Company.

to the problems presented by complex adaptive systems of which life is a part. We will return to this subject in a later section.

Having said all of that, I will try nonetheless to come up with an analytic definition of what life is in the next section and determine whether or not A-Life can find a place in that definition. This definition, due to the reasons I have given above, will be at best only a heuristic tool rather than a complete answer to our question but it will still prove useful in our discussion of A-Life.

Does A-Life Fit our Traditional Concepts of What Life Is?

The first step in answering this question is to sketch a working definition of what 'life' actually means. Life is a powerful yet enigmatic word. We all recognize life when we are confronted with examples of its existence. Very few among us have any problem discerning that the major difference between a statue and a person is that one is alive, and the other is not. We say that plants and animals are alive and that stones and televisions are not. It is easy for us to make these distinctions. But there are things that seem to befuddle our attempts to define accurately the term life. For instance, what are we to make of entities such as viruses? They exhibit some of the properties we associate with life (growth and organization), yet are lacking in many of the qualities of life usually thought to be essential.

Antique Notions of Life

Let's start by briefly reviewing some of the common explanations of what life is. Aristotle is probably the earliest philosopher to turn his attention to the question, what is life? In fact Aristotle is the founder of two different schools of thought on this subject! One theory defines life by referring to various life-functions, and the other, vitalism, maintains that life is given by some sort of vital force or fluid.

The life-functions theory has guided much of the thought on the subject of "what is life, and it would be best for us to discuss its claims and the possible problems with the view. Aristotle suggested that in order for a thing to be considered alive it must posses the following life functions, which I have paraphrased from Fred Feldman in his book published in 1992, *Confrontations with the Reaper*:

> Nutrition: This is the creature's ability to obtain food and absorb it into itself in order to grow. For Aristotle this criteria also includes reproduction which is the creature's ability to produce offspring. One could think of this as requiring that the creature grows both as an individual and as a species. Aristotle feels that this is the prime criteria for determining life and that no thing that is alive is without the nutritive "soul" or ability.
> Sensation: This is split into two subgroups: immediate and mediate sensation. For Aristotle the immediate sensations are those of touch and taste. They are possessed by all animals, but no plants have this ability (Aristotle must not have known about Venus fly traps). Mediate perceptions are things such as sight, hearing, or smell and these are not found in all creatures as, for instance, some immobile creatures do not need the sense of sight and even some mobile deep sea creatures also do not.
> Motion: This is another of the abilities that not all creatures posses but only live things have it.
> Thought: This ability is another specialist category that not all living creatures posses. Yet again for Aristotle, if a thing thinks then it is alive. An example of a thing that only thinks but lacks the other life-functions would be the gods, which are only thinking things.

This list looks good for arguing that artificial life can or maybe already has attained the status of being alive. Simply put: artificial life consists of self-organizing programs that are made up of bits of information. They exist in a world that is composed of other bits of information. These autonomous programs interact with each other and it is claimed that they can *sense* each others presence within the virtual world they inhabit and can distinguish between predator, prey, and potential mates. When they encounter one another they either *eat* the other program if it is prey (thus absorbing its bits of information), or they swap sections of programming with each other if they are of the same *species*, which they use to create new programs that are similar to both *parent* programs but not entirely identical to either. The programs interact in this virtual world and those that are most compatible with a fitness function are allowed to continue into the next generation while the others are erased from the overall program (Rucker, 1993). This is the general process behind the concept of the genetic algorithm.

With this brief overview of artificial life, and our discussions in the first sections of this paper in mind, we can see that as long as one is able to make the conceptual leap that a simulated world on a computer, a world of electrical information exchange, could be conceived of as being analogous to our world of biochemical information exchange, then artificial life can make a claim to being alive at least when the computer is up and running the appropriate program. But even if you can justify this radical claim there are still some fundamental problems with the life-function theory. Of course it is easy to see the basic flaws in this particular view. For instance all of Aristotle's talk of these various "souls" runs afoul of all of the classic problems with dualism. One could easily remove all the talk of souls from this theory with out completely altering it but it would still comes up lacking. Feldman explains some of these other flaws in the life-function theory. For instance if we suggest that all a thing needs to be alive is to posses one or more of the above life-functions then we would be forced to think that "any mechanical device that is capable of setting itself into motion displays this sort of life function . . . [this] seems to imply that alarm clocks, robots of various sorts, automatic lawn sprinkling devices, and the like are all alive" (Feldman, 1992, p. 29). So no matter how tempting this theory is for proving the viability of artificial life, it should not persuade us if it will cause us to have a tinge of moral regret every time we turn off a sprinkler.

There are modern modifications to Aristotle's theories which state that all a thing needs to be alive is that it has the ability to reproduce and is able to pass on genetic variation among its offspring (Feldman, 1992, p. 32). This, according to Feldman, is the view that NASA has come up with when trying to define life in the most abstract way in order to determine if there is the possibility of extraterrestrial life. Limited to these two criteria it would seem that the automatic sprinkler is not a candidate for being a living thing but artificial life passes with flying colors. Unfortunately, Feldman has a good argument that shows that this view is also mistaken. For one thing many creatures are alive but for some reason or another they are currently unable to reproduce (they are too young or too old, etc.). A possible way around this problem is to suggest that the emphasis should be placed on the species rather then the individual, thus changing the argument: a thing is alive if it is a member of a species that has the ability to pass on genetic variation to following generations. Again Feldman counters this argument by suggesting that under this account we would have to say that a preserved dead butterfly is actually alive as it is a member of a species that has the ability to reproduce and pass on genetic variations to its descendant generations (Feldman, 1992, p. 34). Another potential problem is the conceptual ambiguity of the term species.[19] There is currently a debate about just what a species is and it would be unwise to use the concept of species as a criteria for life when it is not well defined itself. Also Feldman suggests that one of the ways we think

---

[19]See, Hull, D. (1989). The Ontological Status of Species as Evolutionary Units. In M. Ruse (Ed.), Philosophy of Biology, New York: Macmillan Publishing Co.

of species is to think of them as groups of living things, so an argument that wishes to link viability with species membership is circular. So even though life-functional theories point out important criteria which we often attribute to living things, they are still lacking some important distinctions that would allow us to, at any time, accurately deduce what is alive from what is not. This is an unfortunate setback for the viability of artificial life, but there are other theories of life that we should look at before we consign artificial life to the heap of things permanently unliving.

Vitalism and DNA-ism

The next theory of life we will look at is the theory of vitalism and its modern descendants. Aristotle seems to also hold to a vitalist theory of life. The chief distinction that separates vitalism from other theories of life is that the vitalists hold that what distinguishes the living from the nonliving is that "living things contain a special substance, which can be called their 'Life'" (Feldman, 1992, p. 40). This substance differs for most vitalists, yet it is always some kind of mysterious, rarefied, fluid similar to something like phlogiston or ether. For instance when we look at how Aristotle answers the question of how a seed in plants, or semen in animals, can grow into an adult member of its species, we see that his answer is that the seed or semen must contain an internal mechanism by which it grows into an adult (Aristotle, in Ruse, 1989, p. 28). Aristotle concludes that this internal mechanism is the soul and that this is the necessary ingredient for life. In fact life and soul are often used synonomusly by Aristotle (In Ruse, 1989, p. 30).

Feldman mentions other Vitalists such as Hans Driesch who posited that life was caused by an imponderable vital fluid which left the body when one dies, much like phlogiston was said to leave a burning thing (Feldman, 1992, p. 41). The major problem with this theory is that it is so unempirical. By definition there is no way to comprehend an "imponderable substance." Feldman notices this problem as well as pointing out two others. The first is that Feldman can not see:

> What difference the vital fluid makes. Animation (whatever that may be) seems to be the crucial factor. So [a] nonvitalistic [definition] is just as plausible as the vitalistic. . . . Each, of course, is hopelessly obscure. (1992, p. 45)

Another problem with the vitalistic position is that even if it turned out that every living thing on earth contains vital fluid, this does not rule out the possibility that there potentially might be living things in other parts of the universe that do not contain this vital fluid (Feldman, 1992, p. 45). So it would seem that vitalism has problems both inductively and deductively.

One may wonder what purpose there is in bashing an old debunked theory like vitalism. The reason is that this theory is not really dead. It has just transformed into new theories, one of which Feldman names DNA-ism.

DNA-ism is the theory that in order for a thing to be alive it must contain some form of DNA or RNA (Feldman, 1992, p. 47). Feldman attacks this theory using a similar tactic to his arguments against vitalism; for instance, he notes that it is possible to remove DNA unharmed from the cells of an animal and put that substance into a test tube, even though the test tube contains DNA it is not considered alive (Feldman, 1992, p. 49). One could say that the DNA has to be properly contained in a cell. In answer to that Feldman suggests that it is not logically imperative to think that even if all living things on earth have DNA and or RNA that some other living thing could exist which has neither. Feldman backs up this argument by quoting the following passage from Francis Crick:

> Of course, elsewhere in the universe life may exist based on other materials. At lower temperatures liquid ammonia might serve as the solvent, though it is not as versatile a solvent as water, which is an exceptionally good one. Instead of carbon, silicon has been suggested. . . . Thus, a form of life based on other materials is not impossible. (In Feldman, 1992, p. 50)

So we find that DNA-ism is untenable, which is good news for artificial life as no form of biological DNA-RNA could be imagined to exist within a computer. In rejecting vitalism and its descendants we have left the door open for artificial life to potentially be considered life but a more positive argument has to be presented. This argument exists and it is found in the science of complexity.

Complexity Theories

Complexity offers a relatively new theory for defining life that is rapidly replacing all of the others we have discussed above. This theory grows out of an older theory presented by Francis Crick which is commonly called genetic informationalism. Genetic informationalism tries to solve the problems of DNA-ism by allowing for multiple possibilities in the media that transmit the genetic information from one generation to another. This theory, while better than DNA-ism, still fails, as it can be used to suggest that some things that we know are dead are actually alive. For instance, Feldman gives the example of a dead tomato plant that happens to have some still viable seeds in it which will grow next season when they fall out of the rotting plant. If you follow genetic informationalism to the letter then you will be forced to say that the tomato plant is actually a living system, and this seems wrong (Feldman, 1992, p. 54).

Feldman ends his discussion by saying that defining life is an enigma and cannot be solved (Feldman, 1992, p 55). Of course Feldman is of the analytic school of philosophy and, as I stated above, there will probably never be an analytic-reductionistic definition of life so it is no wonder that his project does not produce the desired results.

Unbeknownst to Feldman the project of defining life has already taken another step forward. As we discussed earlier, von Neumann's theories involved a complex thought experiment where he proved that it was theoretically possible to build a robotics factory that could replicate itself given the right raw materials and enough time. As we saw in or discussion, his model was grossly impractical in the time and energy required to produce a real working example, but the mere fact that a self-replicating machine could logically exist was an intriguing hypothesis. Remember also that researchers who have taken up this line of thought have focused on von Neumann's mathematical model of this process and have simulated the models on computers:

> These early followers of von Neumann, and the a-life researchers that followed them, encountered something quite remarkable that made their task easier. However counterintuitive it may be, certain natural tendencies or rules of the universe, if you would--seem actually to *encourage* phenomena such as self-reproduction. The efforts of these experimenters became an important tributary to a flow of scientific theory and experimentation: the field of complexity, which bore particular significance to artificial life. (Levey, 1992, p. 30)

Von Neumann felt that the main distinctions between life and nonlife is that life is a system that is grounded in information exchange and complexity (Levey, 1992, p. 30). For a thing to be alive it has to be a system that has the potential to pass on genetic type information. So far this theory is identical to Crick's genetic informationalism but to this hypothesis Von Neumann added a new criterion that requires that the entity must be operating in a significantly complex manner (Levey, 1992, p. 30). This begins to solve the dead tomato problem as the dead parent plant is not currently operating in a complex manner, it is simply decaying; but when the seeds sprout under the right environmental conditions we can use this definition to show that they are alive because they have genetic information in their cells and they are a sufficiently complex system. "The implication for biologists soon became clear. . . . Though there was no mystical *élan vital* that distinguishes life from nonlife, there was something absolutely integral to biological systems that might be considered a sort of life force: complexity" (Levey, 1992, p. 31).

This is an interesting position, but work still needs to be done. For instance, at what point is a system "sufficiently complex"? It seems that we are just replacing one ill-defined term, "life," with another "sufficiently complex," thus changing the readers focus of attention without really advancing our understanding of the world. Often researchers in artificial life rely on the concept of complexity as if it is some kind of magical force. This concept clearly has to be better defined. Yet many researchers have found that chaotic phenomena often produce strange attractors that serve to self-organize immensely complicated functions. Many artificial life researchers feel that the phenomena of evolution is not as aberrant as we might think and they are using the tools of computers and simple programs interacting in complicated ways, to try to prove this thesis empirically.

Theoretical Biology and Complexity

It is interesting to note that the traditional biology consisting of the synthesis of the thoughts of Darwin, Weismann, and Mendel does not have a firm theory about what life is (Stewart, 1992, p. 479). This is most curious as biology is the study of living things and a knowledge of how to differentiate the living from the nonliving might be helpful. The philosopher Marc Lange has written an article for the journal, "Philosophy of Science," where he addresses this very issue (Lange, 1996). It is his feeling that an understanding of what we mean by the term

> life" may perform work in biology, and one goal of biology may be to understand vitality, without one goal of biology being to present a definition (in the reductive sense) of "vitality, any more than one goal of biology is to define "ostrich" or one goal of physics is to define "electron." (For that matter, in order for the distinction between life and none life to perform work in science, an understanding of vitality need not be a goal of science. It is possible that an understanding of vitality is valuable to science as a means, not as an end in itself. (1996, n.p.)

Lange seems to be correct when he makes the above observations. As I mentioned above it does seem to be impossible to obtain a thorough reductionistic definition of life. But it is still important and proper to distinguish between those things that display vitality and those that do not. Lange goes on to remind us that we often use classifications which we do not thoroughly understand in science as heuristic devices to help us form explanations of observed events (Lange, 1996). One of the examples he gives is that in the early days of science before we knew about atomic structure we still had the conception that the ability of something like copper to conduct both heat and electricity probably had something to do with its basic structure. At that time it would be heuristically useful to classify copper as a material that is conductive long before we knew why (Lange, 1996). It is logical then to assume that the concept of life is not chimerical, but rather a useful heuristic category that can be used to classify entities we observe. Recently the ideas of a pair of theoretical biologists who have been working on the problem of defining life for the past few decades have become popular with theorists in A-Life. These thinkers are the team consisting of Humberto Maturana and Francisco Varela. Manturana and Varela feel that they have a reasonable description of what constitutes a living system. Rather than trying to reduce the concept of "life" to a concrete list of life functions, Manturana and Varela suggest that what makes a system living is its peculiar form of organization (Varela, 1979, p. 12). This organization must be what they call autopoietic. The word "autopoiesis was coined by the team in an attempt to find a term that would describe a type of circular organization that they thought was peculiar to living systems, but they wanted to avoid terms that are already used in biology in order to simplify "enormously the task of talking about the organization of the living without falling into the always gaping trap of not saying anything new because the language does not permit it" (Manturana, and Varela, 1972, p. xvii). Broken down, the word simply means *auto*--self, *poiesis*--creation or production. Key to their conception of living systems is the idea that these systems must be autonomous. "A system is autonomous if it can specify its own laws, what is proper to it . . . , the mechanism that makes living beings autonomous systems is autopoiesis (Manturana, and Varela, 1987, p. 48, and 1979, p. 17).

It is not possible to do justice to the autopoietic theory of life in a few paragraphs and a lengthy diversion on this subject is not warranted, but we should come to understand the basics of their theory in order to see its appeal in the area of A-Life research. The technical definition of an autopoietic system is as follows:

> An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components that: (1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produce them; and (2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network. (Manturana, and Varela, 1979, p. 13).

We can see that this definition is conducive to A-Life, as Manturana and Varela have explicitly stated that living systems are autopoietic machines. It is no stretch then to imagine that if what counts in living systems is a specific type of organization, and since it is determined that this organization is a mechanical one, then it would seem that Artificial Life is indeed a possibility. For Maturana and Varela, autopoiesis is a necessary and sufficient condition for living systems (Manturana, and Varela, 1979, p. 17). If it is true that autopoiesis can be realized by machines (they believe that it can only be realized by machines), and since we are capable of understanding the structure of machines and implementing them on alternative media, then A-Life is indeed possible.

It is important to note that Maturana and Varela also admit that:

> Living systems are units of interactions; they exist in an ambiance. From a purely biological point of view they cannot be understood independently of that part of the ambiance with which they interact: the niche; nor can the niche be defined independently of the living system that specifies it. (1979, p. 17).

The theories of Maturana and Varela mostly agree with the comments I made at the beginning of this chapter and I think that they are moving in the right direction and their work is exciting and vital, but I have to question some of their assumptions, the most notable of which is the *a priori* assumption that living systems are some kind of extremely complex machines. It is my opinion that this is an ungrounded assumption. It is trivially obvious that part of what it means to be alive includes the idea that the living thing must be describable at certain levels as some form of bio-mechanical processes. But that does not imply that one can automatically reduce the property of life to certain mechanical organizations. Maturanna and Varela argue that this counter-argument is not cogent because it relies on a naive notion of what a machine is. For them a living machine is very complex, so much so that it is difficult (but not impossible) for us to understand its organization (Manturana, and Varela, 1979, and 1972). I feel that the concept of "sufficiently complex machine does not work philosophically any more than the concept of "living thing. They are both vague and simply reflect the metaphysical presuppositions of their authors and they do not advance our understanding of what life is.

Let us now look at how Mantura and Varela's theories have been used specifically in the pursuit of A-Life. The theoretical biologist John Stewart from the Pasteur Institute in Paris has worked with the ideas of Maturana and Varela and with the basic concepts of complexity. He has offered an interesting argument that attempts to define life. His basic claim is that "living organisms are a subclass of autonomous dynamic systems with the particular property that they continually produce themselves" (Stewart, 1992). Obviously we have to come to an understanding of what Stewart means by the terms "autonomous" and "self-reproduction. Stewart's use of the word "autonomous follows that of Varela in his book, *Principles of Biological Autonomy*. Stewart defines an autonomous system as being neither totally determined by its environment, nor solipsistic or totally independent of its environment (Stewart, 1992, p. 479). In this view a system is autonomous when it is operating in a narrow band somewhere between complete determination and complete chaos; it includes aspects of both ends of the spectrum but is not entirely explained by referring to one aspect over the other. This is very similar to the ideas of Christopher Langton as sketched in Rodger Lewin's book, *Complexity,* where Langton is quoted suggesting that life exists "at the edge of chaos" because this is where nature provides enough structure for life to cling to but also affords the system enough freedom, or autonomy, which allows the organism to change and pursue new survival strategies which both creates and is created by the forces of evolution (Lewin, 1992, p. 186).

If a system has too much structure (like that of a mineral), then the realization of life is impossible due to the system's inability to change in a *frozen* environment. If the environment is too chaotic (like the environments found inside a volcano or superheated gas, for instance) then no complex system can last long enough to propagate itself. The idea that there seems to be only a narrow band of possible niches for life is very reasonable, and seems to be confirmed by the narrow biosphere on our own planet where living systems only appear with in a few miles in either direction of sea level.

Stewart makes an interesting attempt at formalizing a definition of an autonomous system. He feels that "it is possible to describe an autonomous system in terms of formalism which is applicable to simulation models of living systems, and which may hopefully be used as a guideline in attempts to build actual forms of artificial life" (Stewart, 1992, p. 480). In this last quote, Stewart may muddle the distinction between a simulation and that which is simulated; keep this point in mind as we look at his argument.

Here I will attempt to paraphrase Stewart's formal definition for life. As was stated earlier, the first criteria for a living system was that it be an autonomous dynamic system. Stewart claims that a dynamic system is composed of a finite number of elements, $i = 1 \ldots n$, and the state of each of these elements can be defined as x_i, and the global state of the system is given by the state-vector $\mathbf{x}$. Equations of the form:

$$\mathbf{x}' = f(S;\mathbf{x}) \qquad [1]$$

show the dynamics of the system, "where $\mathbf{x}'$ is the time derivative of $\mathbf{x}$, S is the *structure* of the relations between the components, and f is a determinate functional form" (Stewart, 1992, p. 479). Stewart suggests that the system as formally described above escapes being determined by its environment without being totally independent from its environment because of the distinction between the internal state **x** and the structure S (Stewart, 1992, p. 479). The internal state, x, is completely determined by equation [1] since the equation contains no reference to anything which is external to the system (Stewart, 1992, p. 479). But the structure S can be modified by interactions between the system and its environment;

> and this will lead to a modification in the subsequent dynamics and hence to a different internal state **x** than if the interaction had not occurred. Thus, the structure S plays the crucial role of an *interface* between **x** and the environment, both shielding **x** from any direct determination by the environment, and yet at the same time making it possible for the environment to significantly influence **x** (and, a point of some importance for a full development of this theory, for the system to act in return with respect to the environment). We can sum up by saying that the absence of external reference results from the *operational closure* of the system (as expressed in [1]); the simultaneous possibility of significant interaction results from the *structural coupling* between the system and its environment.[20] (1992, p. 479)

This gives us a formal definition for a general autonomous system which seems to be adequate. But more needs to be added before we can claim that it captures anything close to a full definition for living systems. But even though it is not at this time a complete formal representation of life, one may ask whether or not it is the first step in that direction.

So we can see that some thinkers in the field of Complexity, such as Maturana and Varela, have come up with the next step in defining the concept of what life is in a traditionally biological setting. For our purposes we now need to look at the claim that the computer could be an environmental niche as robust as any other niche in which life

[20]Also note that Stewarts sights Varela, 1980, at the end of this explanation.

can exist. If this can be proven then we can imagine that some extrapolation of the formal system described above by Stewart may form the basis for a truly living A-Life program. For this we will look at the proof that Horton Conway, a Cambridge mathematician, provides suggesting that it is conceivable that the functions of nature can be reproduced on a significantly powerful computer.[21]

---

[21]See Levey's discussion of this proof on Pg. 49.

CHAPTER 6

The Game of Life

In the early 1970's the mathematician Horton Conway invented *the Game of Life* (Sigmund, 1993, p. 10). This strange game requires only the nominal presence of a player (whose duties may be automated by a computer), an unbounded "chess board" or the equivalent computer representation of one, and three simple rules (Sigmund, 1993, p. 10). The first rule is that a cell on the board may be either empty or occupied. The second rule is that if a cell is currently empty then it remains so on the next turn unless there are exactly three neighbor cells occupied, in this case the cell will "come alive, and it will be occupied in the next generation. The third rule is that if a cell is occupied then it will remain so whenever two or three of its neighboring cells are occupied; if this is not the case then the cell will be empty and "die in the next generation (Sigmund, 1993, p. 10, and Levey, 1992, p. 49). The "Game of Life" is a very simple cellular automaton. But even though the rules are extremely simple, some very interesting behavior can be generated with the right initial conditions. It is possible to create patterns that remain fixed through many generations, or patterns that move steadily across the game board, or patterns that oscillate between a number of positions.

> The generations unfold like an abstract cartoon. Oscillators pulsate, Gliders wriggle across the screen, some patterns explode, *debris* assembles into new objects, things grow and scatter, collide and vanish, etc. (Sigmund, 1993, p. 12).

Since its invention in the mid-seventies *Life* has attracted many devotees who have studied its operations and cataloged many of the interesting patterns which can be formed through its simple recursive rules, and the researchers have given these patterns descriptive names like *Beehive*, *Blinker*, *Block*, *Glider*, *Pentadeathlon* . . . the list seems endless, and each pattern exhibits a different behavior (Langton, 1989, p. 20). For the most part this research was done in the spirit of fun as Conway presented his findings first in Martin Gardner's "Mathematical Games" column in *Scientific American*.(Langton, 1989, p. 20, and Sigmund, 1993, p. 10).[22] It turns out that all of this activity is much more then just the idle fancy of mathematicians but, as we will see, there may indeed be some payoff for all of this research, especially in the field of A-Life.

At first glance one would think that as long as you know the three simple rules of the game and the initial placement of the *living* cells, one could easily predict the behavior of the system. Well, it turns out that even though many hundreds of researchers have pondered this game either formally or informally, to this date no one has found any regularity to the behavior observed in this game. As an example of this unpredictability let us look at how Karl Sigmund describes how differently the game acts when the initial conditions are just a certain number of *Live* cells in a row:

> Take, for example, a row of occupied cells. What happens? This depends, of course, on the length of the row. But how? Well, three cells in a row form the familiar *Blinker*; four cells in a row change into a *Beehive*; five cells become four *Blinkers*; six cells in a row do a vanishing act which takes 12 generations; seven cells explode and finally form four *Beehives* and four *Blocks*; nine cells in a row end up as four *Blinkers*; 10 cells transform into the *Pentadecathlon*, an amazing pattern which repeats itself after 15 time steps; 11 cells in a row shrink down to a couple of *Blinkers*; 13 cells do the same; 12 cells, however, develop into two *Beehives*; 14 and 15 cells in a row fade off, not without a protracted struggle; 16 cells change into a lively *Traffic Light* built out of eight

---

[22]For an interesting narrative about Conway's invention of the *Life* cellular automaton see. Levy, S. (1992). p. 49.

> *Blinkers*; 17 cells turn into four *Blocks*; 20 cells into two *Blocks*; 18 and 19 cells dissolve like the Cheshire cat. (1993, p. 13).

Conway's motives for inventing and experimenting with the *Game of Life* seem to be similar to the motives that drove Von Neumann in pursuing his Cellular Automaton. As we discussed above, Von Neumann was interested in proving that it was not unthinkable to hold to a mechanistic view of biology. Conway is not after this specifically but he is interested in something similar. What he wants to do is suggest that the *Game of Life* is a model--or metaphor for a universe with a finite number of perfectly known rules of operation (Sigmund, 1993, p. 13). With this premise an argument can be made using the *Game of Life* which concludes that, since the behavior of even a simple universe like that posited in *The Game of Life* can not be predicted, then even if some kind of universal set of rules for our universe could be found it still would not give us the ability to completely predict its behavior from some discrete instant to the next, the universe would still be as unfathomable as before (Sigmund, 1993, p. 13). This result cannot be taken for granted. It has to be shown that *The Game of Life* can produce behavior that is too complex to be completely understood. Conway knew this and he felt that if it could be shown that a pattern could be created that could produce some sort of ordered growth that could go on indefinitely, then he could be assured that his premise was true (Sigmund, 1993, p. 14). After much research and a contest in *Scientific American* Conway believes that he has been able to show that his premises are true. What has been shown is that it is possible to create fairly complex structures called *Glider Guns*, which are patterns that can periodically produce *Gliders*, These gliders are produced like clockwork and continue out to the infinite bounds of the *Life*-plane (Sigmund, 1993, p. 13). At this point Conway's research begins to resemble that of Von Neumann, because after his success at proving that his game was unbounded he then began to look for patterns which he might be able to claim captured the essence of being alive (Sigmund, 1993, p. 15).

As we can recall from our discussions above, Von Neumann's cellular autonoma had twenty-nine states that each cell could be in; others were able to reduce this to seven or eight, but this is still much more complicated then Conway's two state automaton. Another interesting comparison is that Von Neumann's automaton was designed specifically to foster self-reproduction. Conway's was not. In Conway's automaton the ability of certain patterns to propagate was not preconceived and emerged only through experimentation (Sigmund, 1993, p. 28). It is probably safe to assume that Conway's automaton is just about as sleek and unencumbered a simulation of life-like behavior as one is going to get. This makes it perfect for our purposes as it is easy to understand and makes the same sorts of claims that are made for other more sophisticated automata, namely that one has captured, in some form, a suitable amount of the necessary ingredients for life. It is interesting to note that Conway's *Life* automaton turns out to be much more useful then the other self-replicating automata that we have discussed. Both Von Neumann's and Langton's automata have the ability to self-reproduce, which is no mean feat but that seems to be all they are geared up to accomplish. Conway's automaton, on the other hand, can be configured in many different ways so that a skilled *Life* programmer could use this machine to accomplish tasks which need a universal computer to solve. For instance Conway believes that a universal computer made from *The Game of Life* could be employed in the effort to solve any arithmetical problem (Sigmund, 1993, p. 13). The only potential problem here is that Turing has shown that there is no way to predict if the program checking for these solutions will ever stop (Sigmund, 1993, p. 28, and Rucker, 1987, p. 230). Thus we will never know if this procedure will work until we try it. This system is fundamentally unpredictable, which makes it a very interesting sort of machine.

The ability of a *Life* automaton to solve heretofore unsolvable arithmetical problems is an exciting possibility but that is not all that Conway would have us believe his machine is capable of. According to Conway, all that is needed to create a living machine is a universal computer capable of self-reproduction, which can be created by constructing certain patterns on a cellular automaton running a suitable program like *Life*, which in turn is capable of operating like a Turing machine (Sigmund, 1993).

If we swallow Conway's argument then one could conclude that since: (1) A-Life programming can harness the biological-like propagation of discrete bits of information, and (2) on a powerful enough computer a significant level of complexity can be created to rival nature, then it follows that artificial life could be created that could be considered alive in every sense of the word. The weak links in the above premises are of course found in the background metaphysics that is taken for granted in the premises. In premise (1) it is assumed that being biological-like is the same as being biological. And in premise (2) it is assumed that a thing that is nestled in nature, the computer, can create a rival nature of its own. Neither of these two points is obvious and much of the argument rests on these assumptions. Both of these premises seek to blur the distinction between a simulation and that which is simulated. So this brings us to our next question: Is there a significant difference between an extremely accurate simulation of life and the life simulated? This is what we will deal with in the next section.

CHAPTER 7

Cellular Automata as a Biological Environment

We have now read the arguments of some of the leading thinkers in the field of A-Life and we have learned that they believe that life is an organizational structure of autopoietic, autonomous agents interacting in a suitable environment. These organizational structures can be modeled and recreated mechanically. So, it is thought that we may be able to eventually recreate these mechanical structures on some kind of computing machinery. Thus Artificial Life.

One problem with the strong A-Life argument is that it claims that certain things, like cellular automata, or the logic circuits of a digital computer, are capable of creating complex enough environments which can sustain autopoietic machines. This claim seems suspect as these "environments" seem to be many orders of magnitude less complex than nature as we experience it. Is it true that a cellular automaton like *The Game of Life* can capture all the complexity needed to create a living environment?

I have argued above that the best argument currently available for describing living systems is the Maturana/Varela theories. Even though it is still possible to criticize their position, I feel that it is a good working theory. In the Maturana/Varela theory of life we find that the autopoietic organism's interaction with its environment is one of the central aspects of living systems; in fact, it is often difficult to understand an organism without an understanding of its environment, and vice versa.[23] So the diversity and complexity of an artificial environment is the key factor in creating A-Life. The concept that living things are environmentally dependent is not a new

---

[23]See Manturana, H. R., and Varela, F. J. (1972), (1979) & (1987).

idea; its roots in intellectual history can be traced to such thinkers as Spencer and Dewey (Godfrey-Smith, 1994, p. 80). So the arguments claiming that a cellular automaton can become instantiations of living systems need to address the issues raised by theorists who carry on this tradition. Peter Godfrey-Smith, a philosopher at Stanford, writes that, in his view, cellular automata do not fall into this notion of life because a cellular automaton

> is virtually environment-free. The "environment" for a cellular automaton is just the space that it is in, a lattice of cells which can be in various states. The system changes via the local interactions of cells. The system can display complex dynamical behavior in which specific patters or structures are preserved, but it does not generate these patterns as a response to a structured environment. Neither does it respond to potentially disruptive environmental events. The environment does not contain a set of intrinsic patterns which the organic system must adapt to or contend with. Work on cellular automata is one of the more internalist domains within Alife. (1994, p. 84)

This is not to suggest that cellular automata are not useful in studying life. There are many possibilities for cellular automata to be used as extremely powerful models of certain aspects of living systems. But, if one finds the arguments which describe living systems by referencing the living organisms relation to an environment convincing, then one must eliminate cellular automata from the class of living things.

Of course cellular automata are only one of the many breeds of A-Life, we still have to deal with noncellular automata computer simulations and A-Life robotics. In the next section I will present an argument that casts doubt on one of these remaining types but advances the cause of the other.

Gödel's Incompleteness Theorems and Artificial Life

For many decades now it has been claimed that Gödel's two incompleteness theorems preclude the possibility of the development of a true artificial intelligence which could rival the human brain.[24] It is not my purpose to rehash these argument in terms of Cognitive Science. Rather my project here is to look at the two incompleteness theorems and apply them to the field of artificial life. This seems to be a reasonable project as A-Life has often been compared and contrasted to AI, and since there is clearly an overlap between the two studies, criticisms of one might apply to the other (Sober, 1992, and Keeley, 1994). We must also keep in mind that not all criticisms of AI can be automatically applied to A-Life as the two fields of study may be similar but they are not the same (Keeley, 1994).

Gödel himself realized that the incompleteness theorems alone do not preclude the possibility of a machine mind (Wang, 1987, p. 197). In fact there is an interesting argument posed by Rudy Rucker which shows that it is possible to construct a Lucas-style argument using the incompleteness theorems which actually suggests the possibility of creating machine minds (Rucker, 1995, p. 315). Arguments like Rucker's point out the inadequacy of using the incompleteness theorems alone to try to prove the improbability of machine minds. In fact an important part of understanding Gödel's reluctance to accept the project of AI stems from his belief in mathematical realism (Tieszen, 1994, p. 177-201). My purpose here is not to try to convince the reader of the validity of the Penrose-Lucas arguments in cognitive science, but rather to see how a similar argument might be applied to the field of A-Life. I will endeavor to keep my arguments as close to those that Gödel himself might have made if he had been presented with the ideas expressed in strong A-Life. It may turn out that the

[24]See Lucas, J. R. (1961). Minds Machines and Gödel. Philosophy, 36, 112-127. And more recently, Penrose, R. (1995). *Shadows of the Mind.*

incompleteness theorems have no relevance to A-Life but we must take a closer look before we dismiss them out of hand.

A-Life does not start out with the goal of modeling human intelligence, rather it is interested in studying life at a fundamental level comparing and contrasting our knowledge of "*life-as-we-know-it* within the larger picture of *life-as-it-could-be* (Langton, 1987, p. 1). A-Life begins with modest goals. Examples of A-Life projects would range from the modeling of actual biological processes like the life cycles of slime mold (Resnick, 1994, p. 50), to the creation of simple artificial ecosystems like Thomas Ray's *Tierra* program, which is a system entirely resident in a computer and makes few claims to be an accurate representations of life-as-it-is while still claiming to be some new form of synthetic life (Ray, 1991, p. 371). So we can see that at least some of the researchers in the field of A-Life do claim that their creations are (or could be), in a real sense, an actual member of the set of things living (Emmeche, 1994, p. 3).

There are clearly two ways to approach A-Life models: one is to consider them tools for studying the natural world, and the other is to claim that A-Life programs, properly executed, simply are living things (Sober, 1992, p. 749). It seems to me that Gödel's theorems will have little impact on the former claim as it already concedes that A-Life is simply a modeling technique for and not an instantiation of life. Conversely though, Gödel's theorems probably do apply to the much stronger latter claim that A-Life can currently, or will eventually, create artificial living things. This is because the later claim suggests that an artificially constructed reality can capture completely the minimum necessary criteria for the creation of life and, as we will see later, Gödel's theorems can be argued to imply that this may be problematic.

Gödel's Views on Mechanism in Biology

I was spurred in the direction of applying Gödel's theorems to A-Life when I came upon the following passage in Hao Wang's *From Mathematics to Philosophy*, where he is discussing Gödel's views on the relationship between minds and machines:

> Gödel believes that mechanism in biology is a prejudice of our time which will be disproved. In this case one disproval, in Gödel's opinion, will consist in a mathematical theorem to the effect that the formation within geological times of a human body by the laws of physics (or any other laws of a similar nature), starting from a random distribution of elementary particles and the field, is about as unlikely as the separation by chance of the atmosphere into its components.

Mechanistic, or the closely related reductionistic, theories have been part of theoretical biology in one form or another at least since Descartes. I do not want to give the impression that I believe that mechanistic or reductionistic theories form some kind of monolithic doctrine; I realize that there are probably as many different versions of these arguments as there are theorists in the field of biology. Later in this paper I will specify what brand of mechanism and reductionism is employed in strong A-Life arguments.

The various mechanistic and reductionistic theories are historically opposed to the much older and mostly debunked theories of vitalism (Emmeche, 1994). These theories (the former ones more than the latter), along with formism, contextualism, organicism, and a number of other 'sms, mark the major centers of thought in the modern theoretical biology debate (Sattler, 1986).

It occurs to me that A-Life falls curiously on many sides of the debate in the philosophy of biology. For instance A-Life uses the tools of complete mechanization, namely the computer, while at the same time it acknowledges the existence of emergent phenomena (Langton, 1987, p. 81). Neither mechanism nor reductionism is usually thought to be persuaded by arguments appealing to emergence. Facts like this one should make our discussion interesting. It may turn out that A-Life is hopelessly contradictory on this point, or this may provide an escape route for A-Life if we find that Gödel's incompleteness theorems do pose a theoretical roadblock to the mechanistic-reductionism theory in biology, which I will outline later.

What I will attempt to do now is to take a look philosophically at how A-Life relates to a specific form of the mechanistic and reductionistic philosophies of biology and then apply Gödel's incompleteness theorems to that view in the attempt to determine if the project of A-Life can avoid the problems experienced by AI in its encounters with Gödel.

Mechanism and Reductionism in Biology

In this section we will be picking on only two of the above-mentioned world views, namely mechanism, and its closely related theory, reductionism. Furthermore, we will be concerned only with specific formulations of mechanistic and reductionistic theories. This means that we need to be very clear in describing just what we mean by the terms "mechanism" and "reductionism" as they are often used in many different contexts and their meanings can change subtly depending on their use. After we have an adequate understanding of the basic assumptions found in the various mechanistic and reductionistic philosophies of biology, we can then determine if the underlying metaphysical assumptions in A-Life theories should be placed under this heading.

The history of the idea of mechanism is an interesting one but I won't treat you to a retelling of it here. We should understand, however, that it received its greatest boost in popularity in the seventeenth century due to the reaction to the new science of physics by those studying natural philosophy. As we all know, the ability of physics to explain, model, and predict things like planetary orbits astounded the scientific community in the seventeenth century. It occurred to many thinkers of that time that many biological things might also be explainable, modelable, and predictable using the basic laws of physics as they relate to machinery. After all, if one looks at a body it does seem to be a machine of some sort with, for instance, lever actions explaining the workings of muscles and limbs, etc. Descartes was willing to describe all animals as simply machines and even possibly the human body could be reduced to this also, but he was not willing to describe the human mind in this way, Bolder thinkers such as La Mettríe were willing to push the metaphor to the limits and describe humans completely as machines (1748, reprint 1912). The metaphor of the machine or clockwork body is prevalent in this time. From this period of rapid discoveries in physics and mechanics we find those wonderful early A-Life experiments consisting of clockwork animals and people which were built as objects of amusement in the seventeenth and eighteenth centuries.[25]

Over the centuries the mechanistic, and the closely related reductionistic, theories of biology have keep pace with current discoveries in science until today a mechanist can be thought of as one who believes "that an organism is in reality nothing more than a collection of atoms, a simple machine made of organic molecules" (Emmeche, 1994, p. 12). We should note that mechanism, as all theories, changes over time. To be fair we should all realize that the mechanistic theories in biology that Gödel would have been referring to in the quote above have changed and are slightly different today. In the late sixties one could have found many mechanisticly inclined theorists who would claim that it was self-evident that, since biological entities were

[25]See the descriptions of these curiosities in Emmeche, C. (1994). And Langton, C. G. (1987).

physical they had to obey the laws of mechanics, and that meant that living systems were simply matter in motion obeying the laws of classical mechanics (Sattler, 1986, p. 216). But physics has gone much beyond classical mechanics, and many biological mechanists would now agree that it is not possible to accurately describe a living system using only classical mechanics (Sattler, 1986). This is perfectly reasonable. If it is generally accepted that classical mechanics is unsuitable for a complete understanding of non-living matter then how can it be expected to be sufficient for explaining the much more complex actions of living matter? (Sattler, 1986). So it is safe to say that most theorists have outgrown the idea that life can be explained wholly in terms of classical mechanics. Instead what is usually meant is the following:[26]

> 1) Living systems can and/or should be viewed as physico-chemical systems.
>
> 2) Living systems can and/or should be viewed as machines. This kind of mechanism is also known as the machine theory of life.
>
> 3) Living systems can be formally described. There are natural laws which fully describe living systems.

Now it is not necessary for one to hold all three of the above statements true in order to be a biological mechanist. All one has to do is believe at least one of the above statements in order to be referred to as some form of biological mechanist. So a mechanist believes basically that living systems can be *completely* explained by the operation of physical laws on matter such as: classical mechanics, quantum mechanics, complexity theory, etc. Any particular mechanist may think that we do not yet have at our grasp all of the laws we need to understand life, but no mechanist will say that we cannot theoretically discover them in a reasonable amount of time.

---

[26]Paraphrased from Sattler, R. (1986).

Reductionism is the next theory that we have to deal with. Reductionism is related to mechanism in biology in that mechanists wish to *reduce* living systems to a mechanical description. Reductionism is also the name of a more general world view, or scientific strategy. In this world view we are to explain phenomena around us by reducing them to their most basic and simple parts. Once we have an understanding of the components, it is then thought that we have an understanding of the whole. There are many types of reductionist strategies. To help clarify the different categories of reductionism I will turn to the work of John R. Searle. Searle lists five different reductionist strategies in his book, *The Rediscovery of the Mind.* These are: Ontological Reduction, Property Ontological Reduction, Theoretical Reduction, Logical or Definitional Reduction, and Causal Reduction (Searle, 1994). To this list we also should add Epistemological and Methodological reduction.[27] This complexity causes much confusion when one tries to discuss the concept of reductionism, so we should briefly describe each of these strategies.

Ontological Reductionism in theoretical biology occurs when a theory states that a living system is nothing but a collection of physical parts (atoms) being acted upon by the laws of physics. This can be abstracted further by saying that the laws of physics are nothing but a set of formalizable axioms which can be understood separate from physical matter. "Hence, a complete knowledge of the physics and chemistry of life would entail a full understanding of life (Sattler, 1986, p. 218). This concept applies to A-Life theories that promote the belief that, "since we know that it is possible to abstract the logical form of a machine from its physical hardware, it is natural to ask whether it is possible to abstract the logical [form] of an organism from its biochemical wetware (Langton, 1987, p. 21).

[27]See Bonabeau, E. W., and Theraulaz, G. (1994). "Why Do We Need Artificial Life"? In, Artificial Life, 1(number 3). And also (Sattler,1986).

Property Ontological Reduction can occur in theoretical biology and in A-Life when one attempts to describe a property or behavior of a living thing by appealing to low-level phenomena or rules which dictate the behavior. An example of Property Ontological Reduction in A-Life would be if someone claimed that the flocking behavior of birds could be completely reduced, for instance, to the workings of Craig Reynolds' famous *boids* program.[28]

Theoretical, or as it is sometimes called Epistemological, Reductionism is the belief that the theories of one science can be reduced to the theories of another. "[I]n biology the central question of epistemological (theoretical, explanatory) reductionism is whether the laws and theories of biology can be shown to be special cases of the laws and theories of the physical sciences (Dobzhanaky et al. 1977, p. 491, as quoted in Sattler, 1986, p. 221). In A-Life this brand of reductionism appears when the claim is made that the laws of nature might be reducible or capturable in the laws surrounding the information processing of computation.

Logical or Definitional Reductionism "is a relation between words and sentences, where words and sentences referring to one type of entity can be translated without any residue into those referring to another type of entity (Searle, 1994, p. 114). This occurs in A-Life when we use terms usually used in biology to describe events that occur in our computer simulations not metaphorically but descriptively. For instance the words "population, "organism, "fitness, etc., are all used interchangeably in A-Life when describing real and artificial life forms.

Causal Reductionism "is a relation between any two types of things that can have causal powers, where the existence and, *a fortiori*, the causal powers of the reduced entity are shown to be entirely explainable in terms of the causal powers of the

---

[28]Note that I am not claiming that Reynolds himself holds to this idea. To see the *Boids* in action look at the Artificial Life II Video Proceedings, Addison-Wesley, 1992.

reducing phenomena (Searle, 1994, p. 114). This seems to occur in biology when one describes the phenotype as being nothing but the actualization of the genotype. And this occurs in A-Life when we, unarguably, say that the observed behavior of a program is nothing more than the implementation of its program code.

Finally, Methodological Reductionism in biology is the claim that living systems should be studied at their most basic level, either the actual atoms and molecules or their theoretical interactions (Searle, 1994, p. 224). This type of reductionism is similar to theoretical reductionism and clearly this occurs in A-Life when it is suggested that we can gain understanding of the real world by seeing it as the interaction of "information" at either the cellular level or at the level of the patterned interaction of electrons in circuit boards.[29]

So we can see that reductionism is a tool, or strategy, for solving complex problems. There does not seem to be any reason that one has to be a mechanist to use these tools. For instance one could imagine a causal reductionistic vitalist who would believe that life is reducible to the *elan vital* or some other vital essence. Cconversely, one could imagine a mechanist who might believe that living systems can be described metaphorically as machines but that life was not reducible to being only a property of mechanics.

Mechanism and Reductionism in Strong A-Life

As this argument is concerned with the strong A-Life position, I will narrow our discussion of the various reductionistic and mechanistic theories of biology down to the specific types commonly found in strong A-Life claims. The strong argument in

---

[29]See Rucker, R. (1995), for an example of this theory.

A-Life claims that A-Life simulations are, or can be, complete in their formalization of the basic laws describing living systems.

Now, Gödel's incompleteness theorems apply specifically to systems which attempt to completely and consistently axiomatize arithmetic, and generally only to systems which attempt to completely and consistently axiomatize their subject (Nagle and Newman, 1958, p. 100). So if we refer to the three mechanistic theories of life listed above we can begin eliminating the ones that do not apply to the strong A-Life conception of living systems. With this in mind we can eliminate number 1 from the list above as the strong variety of A-Life does not believe that living systems should only be viewed as physico-chemical systems. A-Life is *life-as-it-could-be* not *life-as-we-know-it* (Langton 1989, Pg. 1), and this statement suggests that A-Life is not overly concerned with modeling only physico-chemical systems. Postulates 2 and 3 seem to hold though as strong A-Life theories clearly state that the machine, or formal, theory of life is valid and that simple laws underlie the complex, nonlinear behavior of living systems (Langton, 1989, p. 2).

As far as reductionism is concerned, A-Life theories taken as a whole clearly fit into all the above categories of reductionism.[30] But the strong claim in A-Life clearly relies heavily on property reductionism, causal reductionism, and methodological reductionism. so we can remove the other types of reductionism from our discussion.

Having clarified what we mean by the terms mechanism and reductionism we can now formulate a concise statement of the general beliefs of strong A-Life theories as follows:

1. Living systems are property reducible to the laws described in the theories of complex adaptive systems.

[30]For some discussion of this point see, Bonabeau and Theraulaz, (1994). p. 314.

2. Since a complex adaptive system is casually and methodologically reducible to the mechanistic processes involved in the *computation* of *information* at the fundamental level in nature--it is then conceivable that one could completely formalize all of the laws operating in such a system.

3. These laws can be implemented on the proper type of computing machinery.

Conclusion: A properly conceived A-Life program running in a complex enough computer or robot can correctly be said to be alive.

Now that we have a clearly stated expression of the strong A-Life claim we are at the point where we can apply Gödel's incompleteness theorems to the argument. I believe that Gödel's incompleteness theorems have some bearing on the question of the validity of the strong claim in A-Life since the second premise listed above makes a claim to a level of formal completeness that may be subject to the limitations of formal systems described by Gödel.

Gödel's Incompleteness Theorems Applied to A-Life

In order to show that Gödel's incompleteness theorems have a bearing on A-Life we have to prove that it is necessary for strong A-Life to hold to postulate number 2 as I have stated it above. In order to achieve this I will use Steen Rasmussen's article, *Aspects of Information, Life, Reality, and Physics*, as it does a wonderful job of laying out the logical steps taken in the strong A-Life argument (Rasmussen, 1992, p. 767). Briefly stated his argument goes like this:

1. A universal computer at the Turing machine level can simulate any physical process (Physical Church-Turing thesis).

2. Life is a physical process. Corollary 1: Life can be *simulated* on a universal computer.

3. There exist criteria by which we are able to distinguish living from nonliving objects. Corollary 2, From this postulate it follows that it is possible to determine if some specific computer process is alive or not.

4. An artificial organism must perceive a reality R_2 , which for it is just as real as our "real" reality R_1 is for us (R_1 and R_2 may be the same).

5. R_1 and R_2 have the same ontological status. Using postulate 5 and Corollary 1 we can say that the ontological status of a living process is *independent* of the hardware that carries it. Since R_1 and R_2 are ontologically equal, that is, one is not more real than the other, then actual living systems can be created in a digital computer.

6. It is possible to learn something about the fundamental properties of realities in general, and R_1 in particular, by studying the details of different R_2's. An example of such a property is the physics of a reality.

Postulates 1, 2, and 3 are not completely unproblematic but I will not take that up here, rather we will jump to postulates 4 and 5. In postulate 4, Rasmussen rightly claims that in order for a A-Life program to be alive it has to create an environment that is as real to its inhabitants as nature is to us. In explaining this idea he appeals to a concept called a:

> "Meaning Circuit." The basic idea behind this concept is that the world is a self-synthesized system of existence. On the one hand, physics provides the means for communication (light, sound, etc.). Reality can, thereby, acquire its meaning through a conscious conception of the world, via an organization of the information we get from our senses. On the other hand, physics also gives rise to chemistry and biology, and through them, an observer participation, namely the emergence of life and later the evolution of man. (1992, p. 769)

So what postulate 4 is saying is that the living systems in an artificial reality must have some form of robust interaction and awareness of that reality and this interaction, this "Meaning Circuit," is what makes the artificial reality *Real.* In postulate 5 an interesting jump is made. He claims that "in postulate 4 we argued that a reality obtains its meaning through the existence of an observer(Rasmussen, 1992, p. 770). He then goes on to explain that the artificial reality is a real reality whenever it has a living agent interacting with it. If this is achieved then R_1 and R_2 have equal ontological status (Rasmussen, 1992, p. 770). The problem with this argument so far is that it seems to be circular. It is making the claim that an artificial reality created in the computer is able to capture all of the essential qualities of our reality (R_1 is equal to R_2) as long as living agents are interacting with the system, but the artificial reality must be ontologically equivalent to our reality already in order to produce truly living artificial life forms. In a sense the argument is saying that in order to create artificial life one needs to have artificial life to create the proper artificial environment with the right ontological status--which comes first? I believe that this is a serious flaw in the strong A-Life argument and it may be much more difficult to get around than any of the arguments which will be posed below.

Let's assume that we can get around the circularity of the argument just described. According to postulate number 4, the artificial reality experienced by the artificial life agent must be as real to it as our reality is for us. Using the concept of the Meaning Circuit as described above, it is necessary, in order to capture the essential qualities of the reality we perceive, for an A-Life program to have some form of internal logic equivalent to the physics we perceive in nature so as to provide the artificial organisms with the same kind of meaningful interaction with their world which organisms in our reality experience. This physics can be a simplified version of the one we experience in our reality, but it must be a complete formalization of a certain number of basic physical laws required for the existence of life (Rasmussen, 1992). For instance, there must be some way for the agents and the environment to interact. Since we are programming a computer to invoke this environment, this set of basic physical laws must be one that can be formed into specific statements upon which the program will mechanically deduce the environment and the agents in that environment. We can state this as a postulate:

- There exists a minimum set of formal axioms which can be used to create a complete artificial physics capable of sustaining artificial life.

Now here's the tricky part. One of the main differences between an actual living organism and its potential A-Life counterpart is that the A-Life entity exists in a computer. Also a living creature is presented with the physics of the natural world, whereas an A-Life entity has to have its physics provided by the computer. So in accord with the above postulate, a programmer must code into a computer system the minimum set of formal axioms needed to create a complete artificial physics capable of sustaining artificial life. In order to become a proper artificial physics capable of sustaining life, the program used would have to be able to simulate a reality that is as real to its inhabitants as ours is to us. Now if we hold to a level of mathematical reality as strictly as Gödel does, then concepts like arithmetic are as real an entity as anything else we experience; specifically, a mathematical realist like Gödel believes that our intuitions expressed by mathematics are about "abstract, mind-independent meanings and objects, including transfinite objects (Tieszen, 1994). As we know, Gödel's incompleteness theorems seem to have proven that building a consistent formalized system of proving all arithmetic truths is highly unlikely (Negle and Neuman, 1958, p. 99). Simply put (if that is possible), Gödel's incompleteness theorems suggest that there exists sentences which can be formulated in a specific formal system called Peano-Arithmetic which are true but nonetheless not deducible from the axioms of that system. It follows from this that it is unlikely that we currently have a complete formal system which can grasp the entirety of even simple mathematical systems. This means (as long as you are a mathematical realist) that at least one of the basic qualities of our reality will always be missing from any conceivable artificial reality, namely a consistent and complete formal system of mathematics. This argument tends to make more sense when applied to strong AI claims about intelligent systems understanding concepts.[31] Still, I feel that it has relevance to A-Life for two reasons. The first is that even though the intelligence of a typical postulated A-Life entity is small, it is hoped that greater intelligences will evolve in time from these modest roots. So, if we are to believe that A-Life can eventually evolve higher intelligences we need to know how it can avoid the typical arguments deployed against strong AI claims such as the Gödel argument. Secondly, while one might also ask what possible effect these postulated mathematical realities have on living systems, real or artificial, I believe that it can be argued that some form of mathematical realism is not unthinkable and that this condition of our reality coupled with Gödel's theorems casts doubt on our ability to render an artificial reality which would be equal to our own reality in its ability to sustain life. To illustrate this idea let us look briefly at a quote from John von Neumann regarding mathematics and AI:

> When we talk mathematics, we may be discussing a *secondary* language, built on the *primary* language truly used by the central nervous system. Thus the outward forms of *our* mathematics are not absolutely relevant from the point of view of evaluating what the mathematical or logical language *truly* used by the central nervous system is. (Quoted by Weizenbaum, 1976).

---

[31] See Tieszen, 1994, for a more complete argument as it concerns AI.

It seems that one could broaden the scope of von Neumann's observation from the specifics of a living central nervous system to life in general with out harming the intent of the original comment. I feel that this is the position that a mathematical realist like Gödel would take because a mathematical realist would believe that there exist mathematical realities which are the foundations of the reality we experience and that these realities are described by concepts like Peano-Arithmetic, but that these realities are uncapturable in any complete way by entirely mechanical processes. Thus it would seem that it is impossible to *completely* formalize an artificial reality that is equal to the one we experience, so A-Life systems entirely resident in a computer must remain, for anyone persuaded by the mathematical realism posited by Gödel, a science which can only be capable of potentially creating extremely robust *simulations* of living systems but never one that can become a complete instantiation of a living system.

Objections

The argument that I have presented above is admittedly brief. In a short paper such as this it is hard to adequately defend a theory that makes use of Gödel's theorems as seen from the perspective of his mathematical realism. Both of these subjects would take up the better part of a book to thoroughly explain. My purpose here is only to open a discussion of this topic in the hope that others agree that it is a worthwhile subject for further study. In fact I hope to collect many objections to the argument so that I can attempt to answer them later in a more thorough way.

Still it would be helpful here to look at the most common objection that I have received to this argument and attempt to begin a counter argument.

One objection we can think of is this: Our reality (R_1) is a reality in which the incompleteness theorems hold. So why does it matter that the incompleteness theorems hold in an artificial reality (R_2)? All the above argument has accomplished is to point out that Gödel's theorems are valid in both R_1 and R_2 . Also, computers already do some amazing things, none of which requires the strict formal completeness and consistency that Gödel is worried about in his famous theorems.

It is true that the incompleteness theorems hold to our perceived reality and that they point to a fundamental limit to our ability to formalize all of our mathematical intuitions. I do not believe that Gödel meant to suggest that mathematics as a separate entity is fundamentally incomplete. Rather his theorems prove that our understanding of that mental object known as mathematics cannot be completely and consistently mechanized. So what I am saying is: given Gödel's mathematical realism, the incompleteness theorems suggest that it is not possible to capture this one aspect of our reality in any artificial reality. If one assumes that our universe is infinite, "then it embodies the full set of natural numbers, so Gödel's Theorem seems to say that for any given finite theory of the universe, there are certain facts having to do with sets of physical objects that cannot be proved by the theory (Rucker, 1995, p. 141). Now any A-Life program that is attempting to create an environment which is capable of sustaining life and which is entirely separate from our own environment is attempting to capture the sufficient conditions which makes life possible here. I am claiming that Gödel's Theorem suggests that any such program might be missing important essential portions of our reality, namely its fundamental mathematical reality, so that the artificial reality (R_2) would not be ontologically equal to our reality (R_1), and since this is a requirement for creating truly living artificial life entities, the artificial reality could not sustain life.

So What?

Now I will try to mitigate some of the consequences of the above argument and suggest ways that A-Life can avoid the argument, or change to accommodate it.

We should not feel that A-Life is diminished if it proves to be impossible to synthesize living systems in the manner described above. A-Life in its so-called "weak" form is still a challenging new science which promises to completely alter the way we practice the study of biology by giving us powerful new tools and metaphors for looking at and discussing living systems (Emmeche, 1994, p. 156). Secondly, the arguments given above only apply to A-Life experiments completely carried out within a computer.

When we look at the argument above we can see that all it suggests is that there is not a complete one-to-one correspondence between nature and a simulated nature. Remember that the artificial organism must perceive a reality that is as real to it as our reality is to us (Rasmusen, 1992, p. 769). Since there may be some problem with a simulated reality, then that problem can be solved by allowing the artificial organism to interact with our reality. This can be done through robotics.

In this scheme the robotic artificial organisms are operating in an unarguably real environment. If a way could be found to give the robots complex adaptive behavior and self-reproduction then we might be on our way to creating true artificial life. It may be possible, but certainly not easy, to evolve living organisms from robotics.

CHAPTER 8

Conclusions

In the last chapter we have seen that, due to a specific interpretation of the implications of Gödel's incompleteness theorems, it may not be possible to create a truly living system which is entirely resident in a computer. We were not able to advance very far Gödel's claim that mechanism in biology can be disproved mathematically. We have only proven that life may not be reducible to a certain type of mechanical implementation on a computer. This modest result may lead to a more complete refutation of mechanism but that question is left open for now. It may be that studies in A-Life itself will lead to the mathematical proof that Gödel postulated in the quote above.

The value of this finding is not to discourage certain types of research in A-Life, but rather to help move us in a direction where we can more clearly define the results of that research. In fact, since one of the above arguments rests on the assumption that the universe is infinite and that some form of mathematical realism is true, if we are someday able to complete the goal advanced in strong A-Life it would seem to cast doubt on the validity of the assumptions made above. So succeed or fail, A-Life gives us much to ponder.

It may be that A-Life is still a long way off from capturing completely the answer to the question, "What is life?" It may be that this question is unanswerable or the wrong question to ask. But every attempt at answering that question, from the modest attempts in A-Life at the explication of life, to the extreme attempts in strong

A-Life to synthesize life, helps us move closer to an understanding of the world we find ourselves in.

We also brought into question the ability of a cellular automaton to fulfill the minimum requirements for living systems due to its inability to create an environment suitable as a biological niche. Next we need to look at what A-Life really is so that we can make a proper judgment of its claims and abilities.

Simulations and Simulacra

Now we have to answer the question, What is Artificial Life? Is it a new kind of life form? I feel that the arguments presented in the last chapter suggest that there may be some problems with the project of creating living systems which are entirely resident in a computer. But we have also seen that when an A-Life project allows the machine to interact with our environment then the possibility for creating life seems to become more of a potentiality. Still, there are plenty of obstacles to the eventuality of realizing the strong A-Life argument. It has taken a staggeringly long time for life to develop as we know it. Even if we could speed up the process a little by quickly moving through generations of autopoietic robots every year we would still be confronted with many decades or even centuries of run time for such an experiment to begin to show unambiguous signs of life. So we should answer this question by stating what Artificial Life is today.

Claus Emmeche has concluded that A-Life is a postmodern science (Emmeche, 1994, p. 158). He has consciously linked the science of A-Life to a much larger cultural movement "within philosophy, literature, art, politics, social sciences, and in social life generally, as it unfolds in our industrialized, information-dominated, media-oriented society" (Emmeche, 1994, p. 158). He wants to portray A-Life as part of the process our modern societies find themselves in where we seem to be creating a separate mediated reality that distances us from the real and the material. In his argument, Emmeche uses the technical term "simulacrum. This term is taken from the French sociologist and philosopher Jean Baudrillard (Baudrillard, 1994). Simulations attempt to blur the distinctions between real and ideal, subject and object model and modeled; when this movement is taken to its extreme you have the phenomenon Baudrillard calls a simulacrum (Baudrillard, 1994, p. 31). As a heuristic device one can think of a simulacrum as being a simulation of a simulation. The term "simulacrum" has a specific meaning in the work of Baudrillard in which he describes various orders of simulacra ranked on their distance from a base reality. Emmeche feels that this is exactly what is happening in A-Life:

> It is characteristic of the simulations of life produced by A-lifers and cellular-automata specialists that these simulations do not derive from any natural domain, let alone an experimental physical system. The reference to substance, if not lost, is then left to others. Instead, the models help to produce a fictive reality, where the distinction between copy and original, description and reality, is meaningless. The model no longer seeks to legitimate itself with any requirement for truth or accuracy. It creates a simulacrum, its own universe, where the criteria for computational sophistication replaces truth, and only have meaning within the artificial reality itself. (Emmeche, 1994, p. 160)

This view of A-Life can begin to make one uneasy about its work and implications. How can we trust a science so removed from the reality we perceive? Many theoretical biologists have been bothered by the simulacral quality of A-Life, in fact the biologist John Maynard Smith has called A-Life "fact-free biology" (Horgan, 1996).

I believe that it is true that A-Life can be described as a simulacrum but that is not so damning a charge. If one reads Baudrillard one will see that nothing that is socially produced is free from being described as some form of simulacrum. There is a large and important literature on the idea that science is not immune from social coloration of its theories.[32] If this view is correct then even biology itself is guilty of what we are criticizing A-Life for. Emmeche feels that A-Life is completely legitimate as a field of study for the very reason that it begins to call into question the assumptions we take for granted in biology; for Emmeche (1994), A-Life is a postmodern science that is helping to deconstruct traditional biology.

> From this perspective, artificial life must be seen as a sign of the emergence of a new set of postmodern sciences, postmodern because they have renounced or strongly downgraded the challenge of providing us with a truthful image of one real world, and instead have taken on the mission of exploring the possibilities and impossibilities of virtual worlds. It is a case of *modal* sciences, passing freely between necessity and possibility. Science becomes the art of the possible because the interesting questions are no longer how the world is, but how it could be, and how we can most effectively create other universes--given this or that set of computational resources. (p. 161)

Thus A-Life can be seen as a science that is still trying to define itself and its goals, but it is a vital science in that it will help us renegotiate the current paradigms that we labor under in biology.

Metaphorical Merry-Go-Round

We have briefly talked about the history of the machine metaphor for life in the above chapters. We have seen that La Mettrie published a bold theory in 1748 that pushed the concept of man as a machine into the public debate; no longer could the distinction between man and animals championed by Descartes go unchallenged

---

[32]See the work of Paul Feyerabend, Sandra Harding, Thomas Khun, Francois Lyotard, Sharon Traweek, etc., for more complete arguments.

(Asendorf, 1993, p. 7). Thus the last holdout for describing all living things as machines was broken down. It is true that one can make a list of all known life functions and show that a mechanical process can be described that matches it. I would agree that part of what it means to be alive is describable mechanically; it would be foolish to argue otherwise. But my question is whether it is not too soon to pass the judgment that we will soon completely understand what life is. Each major scientific paradigm seeks to define life in terms of the dominant theory of its time. We can find theories that life is a mystical substance, that it is based on the laws of thermodynamics, that it is reducible to the flow of information such as that in a computer. It seems likely that the definition of life will be further expanded when we reach a new paradigm. We can see that in the eighteenth century for La Mettrie the body was a mechanical object; in the nineteenth century the machine was transformed into a bodily object and the idea of the living machine became credible (Asendorf, 1993, pp. 44-45). In the twentieth century the machine (and therefore the body) has been abstracted to the concept of information. How can we say that the vague cybernetical concept of "information" is any more cogent then any vitalist theory when both rely on an intangible theoretical entity to do the work of the theory?

So we seem to be running in theoretical circles in regards to our understanding of life.

Is Life a Mechanical Process?

I cannot avoid this question any longer. I am reluctant to answer it as I feel that it is a malformed question. The concept of "mechanical" is too vague and caries too many naive suppositions. If by machine one means a simple classically defined apparatus which is predetermined to follow a preprogrammed function, then no, life is not a machine for the various reasons outlined in the chapters above, namely that a classical machine does not capture all of what we observe in nature. If one means a machine that is autopoietic and interacting with a complex dynamic environment, then, yes, at some level life is a mechanical process. But this conclusion does not carry with it the problems of classical mechanics. Complex dynamic systems are unpredictable and autopoietic machines have some ability to modify their behaviors so one does not have to give up hope for a definition of life that is rich and inspiring in an almost religious way. But I do not feel that we have reached the end of the story. I feel that one can describe a continuum where on one end we have non-life and the other life; things like viruses and A-Life are somewhere in the middle of this continuum and we are near the end. It may be possible for entities like A-Life to evolve mechanically up that scale and eventually reach the status of living, but at that point they will no longer be describable as machines.

Works Cited

Aristotle. (1989). The Generation of Animals. In M. Ruse (Ed.), Philosophy of Biology, New York: Macmillan Publishing Co.

Asendorf, C. (1993). Batteries of Life, on the History of Things and Their Perception in Modernity (Don Reneau, Trans.). Berkeley: University of California Press.

Baudrillard, J. (1994). Simulara and Simulation (Shila Faria Glaser, Trans.). Ann Arbor, Michigan: The University of Michigan Press.

Bedeau, M. A. (1992). Philosophical Aspects of Artificial Life. Paper presented at Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.

Bonabeau, E. W., & Theraulaz, G. (1994). Why Do We Need Artificial Life? Artificial Life, 1(3).

Dawkins, R. (1995,). Revolutionary Evolutionist. Wired, 120-186.

Dennett, D. (1994). Artificial Life as Philosophy. Artificial Life, 1(Number 3).

Dreyfus, H. L.(1992). Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branch-Point. In M. A. Boden (Ed.), The Philosophy of Artificial Intelligence. Oxford: Oxford University Press.

Emmeche, C. (1992,). Life as an Abstract Phenomenon: Is Artificial Life Possible? Paper presented at Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.

Emmeche, C. (1994). The Garden in the Machine: The Emerging Science of Artificial Life. Princeton, New Jersey: Princeton University Press.

Feldman, F. (1992). Confrontations with the Reaper: A Philosophical Study of the Nature and Value of Death. Oxford: Oxford University Press.

Feyerabend, P. (1988). How to Defend Society Against Science. Buffalo, NY: Prometheus Books.

Freedman, D. H. (1994). Brainmakers. New York: Simon & Schuster.

Godfrey-Smith, P. (1994). Spencer and Dewey on Life and Mind. Paper presented at Artificial Life IV, Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems, MIT.

Harding, S. (1991). Whose Science? Whose Knowledge? Ithaca, NY: Cornell University Press.

Horgan, J. (1996). Round table discussion on The Talk of the Nation [Radio Broadcast]: Public Radio International.

Hull, D. (1989). The Ontological Status of Species as Evolutionary Units. In M. Ruse (Ed.), Philosophy of Biology, New York: Macmillan Publishing Co.

Keeley, B. L. (1994). Against the Global Replacement: On the Application of the Philosophy of Artificial Intelligence to Artificial Life. Paper presented at Artificial Life III, Sante Fe Institute, Santa Fe, NM.

Koza, J. R. (1991). Genetic Evolution and Co-Evolution of Computer Programs. Paper presented at Artificial Life II: A Proceedings Volume in the Sante Fe Institute Studies in the Sciences of Complexity, Sante Fe, NM..

La Mettríe, Julíen Offray de. (Gertrude C. Bussey Ed.). (1748 [reprint 1912]). L'Homme Machine. La Salle, IL: Open Court.

Lange, M. (1996). Life, 'Artificial Life,' and Scientific Explanation. Philosophy of Science, forthcoming.

Langton, C. G. (1987, 1989). Artificial Life. Paper presented at Artificial Life: The proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems, Sante Fe Institute, Los Alamos, New Mexico.

Langton, C. G. (1993-94). Editor's Introduction. Artificial Life, 1(1 & 2).

Langton, C. G. (1995). Artificial Life Lecture given at Hewlett-Packard Central Offices in Palo Alto, CA.

Latham, W., & Tod, S. (1992). Artificial Life or Surreal Art? Paper presented at the Towards a Practice of Autonomous Systems; Proceedings of the First European Conference on Artificial Life.

Latham, W., & Tod, S. (1992). Evolutionary Art and Computers. San Diego: Academic Press.

Levy, S. (1992). Artificial Life: A Report From the Frontier Where Computers Meet Biology. New York: Vintage Books Edition.

Lewin, R. (1992). Complexity: Life at the Edge of Chaos. New York: Macmillan Publishing Company.

Lucas, J. R. (1961). Minds Machines and Gödel. Philosophy, 36, 112-127.

Manturana, H. R., & Varela, F. J. (1987). The Tree of Knowledge. Boston: Shambala New Science Library.

Manturana, H. R., & Varela, F. J. (1972). Autopoiesis and Cognition: The realization of the Living. Boston: D. Reidel Publishing Company.

Manturana, H. R., & Varela, F. J. (1979). Principles of Biological Autonomy. New York: Elsevier North Holland, Inc.

Nagle, E., & Newman, J. (1958). Godel's Proof. New York: New York University Press.

Penrose, L. S. (1992). Automatic Mechanical Self-Replication [Video proceedings for Artificial Life II]. Redwood City, CA: Addison-Wesley.

Penrose, R. 1994 . Shadows of the Mind: a search for the missing science of consciousness. Oxford, New York: Oxford University Press.

Rasmussen, S. (1992). Aspects of Information, Life, Reality, and Physics. Paper presented at the Artificial Life II, Santa Fe Institute, Santa Fe, NM.

Ray, T. (1991). An Approach to the Synthesis of Life. Paper presented at Artificial Life II, Santa Fe, NM.

Resnick, M. (1994). Turtles, Termites, and Traffic Jams. Cambridge, MA: MIT Press.

Rucker, R. (1987). Mind Tools. Boston: Houghton Mifflin Company.

Rucker, R. (1989). Cellular Automata Laboratory (Version 1). Sausalito, CA: Autodesk.

Rucker, R. (1993). Artificial Life Lab [Computer Program and Manual]: Waite Group Press.

Rucker, R. (1995). Infinity and the Mind: The Science and Philosophy of the Infinite. Princeton, NJ: Princeton University Press.

Sattler, R. (1986). Bio-Philosophy: Analytic and Holistic Perspectives. Berlin: Springer Verlag.

Searle, J. R. (1994). The Rediscovery of the Mind. Cambridge, MA: MIT Press.

Sigmund, K. (1993). Games of Life: Explorations in Ecology Evolution and Behavior. Oxford: Oxford University Press.

Simon, H. (1994, July). Interview, Herbert A. Simon. Omni, 72-89.

Sober, E. (1992). Learning from Functionalisim--Prospects for Strong Artificial Life. Paper presented at Artificial Life II, Santa Fe Institute.

Stewart, J. (1992). Life=Cognition. Paper presented at the Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.

Tieszen, R. (1994). Mathematical Realisim and Godel's Incompleteness Theorems. Philosophia Mathematica, 2(3), 177-201.

Turing, A. M. (1992). Computing Machinery and Intelligence. In M. A. Boden (Ed.), The Philosophy of Artificial Intelligence. Oxford: Oxford University Press.

Varela, F. J. (1979). Principles of Biological Autonomy. New York; Oxford: Elsevier North Holland.

Wang, H. (1974). From Mathematics to Philosophy. New York: Rutledge, Kegan & Paul.

Wang, H. (1987). Reflections on Kurt Gödel. Cambridge, MA: MIT Press.

Weizenbaum, J. (1976). Computer Power and Human Reason: From Judgment to Calculation. San Francisco: W. H. Freeman and Co.

Winner, L. (1986). The Whale and the Reactor. Chicago: University of Chicago Press.