GÖDEL'S INCOMPLETENESS THEOREMS AND ARTIFICIAL LIFE

John P. Sullins III, San Jose State University

In this paper I discuss whether Gödel's incompleteness theorems have any implications for studies in Artificial Life (AL). Since Gödel's incompleteness theorems have been used to argue against certain mechanistic theories of the mind, it seems natural to attempt to apply the theorems to certain strong mechanistic arguments postulated by some AL theorists.

We find that an argument using the incompleteness theorems can not be constructed that will block the hard AL claim, specifically in the field of robotics. However, we will see that the beginnings of an argument casting doubt on our ability to create living systems entirely resident in a computer environment might be suggested by looking at the incompleteness theorems from the point of view of Gödel's belief in mathematical realism.

1. INTRODUCTION

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.[1] 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 AL. This seems to be a reasonable project as AL has often been compared and contrasted to AI (Sober, 1992; Keeley, 1994); and since there is clearly an overlap between the two studies, criticisms of one might apply to the other. We must also keep in mind that not all criticisms of AI can be automatically applied to AL; 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, pg. 197). In fact there is an interesting argument posed by Rudy Rucker where he 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, 1983, pp. 315-317). 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 (see Tieszen, 1994, for a full discussion of this point). My purpose here is not to try to convince anyone 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 AL. 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 AL. It may turn out that the incompleteness theorems have no relevance to AL, but we must take a closer look before we dismiss them out of hand.

AL 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, pp. 1). AL begins with modest goals. Examples of AL 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, a system entirely resident in a computer which 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, 1992, p. 371). So we can see that at least some of the researchers in the field of AL do claim that these creations are (or could be) in a real sense an actual member of the set of things living (see Emmeche, 1994, p. 3).

There are clearly two ways to approach AL models: one is to consider them tools for studying the natural world, and the other is to claim that AL 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 AL 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 AL can currently, or will eventually, create artificial living things. This is because the later claim suggests that an artificially constructed reality can completely capture 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.

2. GÖDEL'S VIEWS ON MECHANISM IN BIOLOGY

I was spurred in the direction of applying Gödel's theorems to AL 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 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 which brand of mechanism and reductionism is employed in strong AL arguments.

The various mechanistic and reductionistic theories are historically opposed to the much older and mostly debunked theories of vitalism (see Emmeche, 1991). These theories (the former more than the latter), along with formism, contextualism, organicism, and a number of other "isms" mark the major centers of thought in the modern theoretical biology debate (see Sattler, 1986).

It occurs to me that AL falls curiously on many sides of these debates in the philosophy of biology. For instance AL 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 should make our discussion interesting. It may turn out that AL is hopelessly contradictory on this point, or it may provide an escape route for AL if we find that Gödel's incompleteness theorems do pose a theoretical road block to the mechanistic-reductionistic theories in biology which I will outline later.

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

3. MECHANISM AND REDUCTIONISM IN BIOLOGY

In this paper I will be discussing only two of the above mentioned world views, namely, mechanism and its closely related theory reductionism. Furthermore, I 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 AL theories should be placed under this heading.

The history of the idea of mechanism is an interesting one, but I will not retell it here. We should understand, however, that it received its greatest boost in popularity in the seventeenth century as a reaction to the new science of physics on the part of those studying natural philosophy. As we all know, the capacity 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 among other things. Descartes was willing to describe all animals as simple machines; possibly even the human body could be reduced this way. But he was not willing to so describe the human mind. Bolder thinkers such as De la Mettrie (1748) were willing to push the metaphor to the limits, describing humans completely as machines. The metaphor of the machine or "clockwork body" is still prevalent today. In this period of rapid discoveries in physics and mechanics we find wonderful early AL experiments consisting of clockwork people and animals which were built as objects of amusement in the seventeenth and eighteenth centuries (see Emmeche, 1991, and Langton, 1987).

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, 1991, p. 12). We should note that mechanism, like all theories, changes over time. To be fair, we should 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 mechanistically inclined theorists who would claim that it was self-evident that, since biological entities are physical they must 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 far 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, p. 216). This is perfectly reasonable. If it is generally accepted that classical mechanics is unsuitable for a complete understanding of nonliving 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 (paraphrased from Sattler, 1986):

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 in order to be a biological mechanist. All one has to do is believe at least one of the above statements. So a mechanist believes, basically, that living systems can be *completely* explained by the operation of the physical laws of matter, such as classical mechanics, quantum mechanics, complexity theory, etc. Any particular mechanist may think that we do not yet have within 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.

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 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, 1992). And to this list we should also add Epistemological and Methodological Reduction (see Bonabeau and Theraulaz, 1994, and Sattler, 1986). 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 AL 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).

Property ontological reduction can occur in theoretical biology and in AL 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 AL would be if some one claimed that the flocking behavior of birds could be completely reduced, for instance, to the workings of Craig Reynolds's famous *boids* program.[2]

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. "In 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 AL 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 AL 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 AL 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 reducing phenomena" (Searle, 1994, p. 114). This seems to occur in biology when one describes phenotype as being nothing but the actualization of the genotype. And this occurs in AL when we say, unarguably, 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 (Sattler, 1986, p. 224). Clearly this occurs in AL 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 (see Rucker, 1987, for an example).

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. And, conversely, 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.

4. MECHANISM AND REDUCTIONISM IN STRONG AL

As this paper is concerned with strong AL arguments, I will narrow down our discussion of the various reductionistic and mechanistic theories of biology to the specific types commonly found in strong AL claims. The strong argument claims that AL simulations are, or can be, complete in their formalization of the basic laws describing living systems.

Now since 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, Braithwaite, 1962, p. 1). 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 AL conception of living systems. With this in mind we can eliminate number 1 from the list above, as the strong variety of AL does not believe that living systems should only be viewed as physico-chemical systems. AL is *life-as-it-could-be*, not *life-as-we-know-it* (Langton, 1989, p. 1), and this statement suggests that AL is not overly concerned with modeling only physico-chemical systems. Postulates 2 and 3 seem to hold, though, as strong AL 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, AL theories taken as a whole clearly fit into all the above categories of reductionism (for some discussion of this point see Bonabeau and Theraulaz, 1994, p. 314). But the strong claim in AL 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 mechanisim and reductionism, we can now formulate a concise statement of the general beliefs of strong AL theories as follows:

1. Living systems are properly reducible to the laws described in the theories of complex adaptive systems.
2. Since a complex adaptive system is causally 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 AL 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 AL 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 AL since the second premise just listed makes a claim to a level of formal completeness that may be subject to the limitations of formal systems described by Gödel.

5. GÖDEL'S INCOMPLETENESS THEOREMS APPLIED TO AL

In order to show that Gödel's incompleteness theorems have a bearing on AL, we have to prove that it is necessary for strong AL to hold to postulate number 2 as I have stated it above. In order to achieve this I will use Steen Rasmussen's (1992) article, "Aspects of Information, Life, Reality, and Physics" (p. 767), as it does a wonderful job of laying out the logical steps taken in the strong AL argument. 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, Hence life can be *simulated* on a universal computer.
3. There exist criteria by which we are able to distinguish living from non-living 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 an AL 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 (Rasmussen, 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 already be ontologically equivalent to our reality in order to produce truly living artificial life forms. So 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 AL argument, and it may be much more difficult to get around than any of the arguments which will be posed below.

Let us assume that we can get around the circularity of the argument just described. According to postulate 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 AL 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 (Rasmussen, 1992, p. 769), but it must be a complete formalization of a certain number of basic physical laws required for the existence of life. 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, then this set of basic physical laws must be one that can be formed into specific statements in 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 is the tricky part. One of the main differences between an actual living organism and its potential AL counterpart is that the AL entity exists in a computer. Also a living creature is presented with the physics of the natural world, where an AL 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 (Gödel, 1962, p. 77, Nagle and Newman, 1958, p. 99). Simply put (if that is possible), Gödel's incompleteness theorems suggest that there exist 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 complete formal system of mathematics. This argument tends to make more sense when applied to strong AI claims about intelligent systems understanding concepts (see Tieszen, 1994, for a more complete argument as it concerns AI).

Still, I feel that it has relevance to AL for two reasons. The first is that even though the intelligence of a typical postulated AL entity is small, it is hoped that greater intelligences will evolve in time from these modest roots. So, if we are to believe that AL 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).

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 without 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 AL 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.

6. 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.

Those to whom I have shown earlier drafts of this paper usually point out an objection similar to 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 in 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 can not be completely and consistently mechanized. So what I am saying is this: 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 can not be proved by the theory" (Rucker, 1982, p. 141). Now any AL program that is attempting to entirely create an environment separate from our own which is capable of sustaining life is attempting to capture the sufficient conditions which make life possible here. I am claiming that Gödel's theorem suggests that any such program might be missing an important essential portion 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.

7. SO WHAT

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

We should not feel that AL is diminished if it proves to be impossible to synthesize living systems in the manner described above. AL 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 argument given above only applies to AL 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 (Rasmussen, 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 robots.

8. CONCLUSION

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 disproven 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 AL 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 AL, 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 AL it would seem to cast doubt on the validity of the assumptions made above. So succeed or fail AL gives us much to ponder.

It may be that AL is still a long way 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 modest attempts in AL at the explication of life, to extreme attempts in strong AL to synthesize life, helps us move closer to an understanding of the world we find ourselves in.

ACKNOWLEDGMENT

I would like to thank the staff and faculty of the San Jose State University philosophy department for their support of my studies. I am also indebted to Dr. S. D. N. Cook for his critique and support of this project, Dr. R. Rucker for his scathing criticisms, and Dr. R. Tieszen for his comments on earlier drafts of this paper. This work has been partially supported through a grant from the National Science Foundation.

REFERENCES

Bonabeau, E. W., and G. Theraulaz. 1994. "Why Do We Need Artificial Life?" *Artificial Life* 1:3 (Spring).

Emmeche, C. 1994. *The Garden in the Machine: The Emerging Science of Artificial Life.* Princeton, N.J.: Princeton University Press.

Keeley, B. L. 1994. "Against the Global Replacement: On the Application of the Philosophy of Artificial Intelligence to Artificial Life." In *Artificial Life III,* ed. C. G. Langton. SFI Studies in the Science of Complexity, vol. 17. Reading, Mass.: Addison-Wesley.

Langton, C. G. 1987. "Artificial Life." In *Artificial Life,* ed. C. G. Langton. SFI Studies in the

Science of Complexity, vol. 6. Reading, Mass.: Addison-Wesley.

de la Mettrie, Julien Offray. "L'Homme Machine, 1748." In *Man a Machine*, ed. G. C. Bussey. La Salle, Ill.: Open Court.

Gödel, K. 1962. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems,* ed. B. Meltzer. New York: Basic Books.

Lucas, J. R. 1961. "Minds, Machines and Gödel." *Philosophy,* 36:112-127.

Nagel, E., and J. R. Newman. 1958. *Gödel's Proof.* New York: New York University Press.

Penrose, R. 1994. *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New York: Oxford University Press.

Rasmussen, S. 1992. "Aspects of Information, Life, Reality, and Physics." In *Artificial Life II,* ed. C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen. SFI Studies in the Science of Complexity, vol. 10. Reading, Mass.: Addison- Wesley.

Resnick, M. 1994. *Turtles, Termites, and Traffic Jams*. Cambridge, Mass.: MIT Press.

Rucker, R. 1995. *Infinity and the Mind: The Science and Philosophy of the Infinite*. Princeton, N.J.: Princeton University Press.

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

Searle, J. R. 1994. *The Rediscovery of the Mind*. Cambridge, Mass.: MIT Press.

Sober, E. 1992. "Learning from Functionalism—Prospects for Strong Artificial Life." In *Artificial Life II,* ed. C. G. Langton, C. E. Taylor, J. D. Farmer, and S. Rasmussen. SFI Studies in the Science of Complexity, vol. 10. Reading, Mass.: Addison-Wesley.

Tieszen, R. 1994. "Mathematical Realism and Gödel's Incompleteness Theorems." *Philosophia Mathematica* (3), 2:177-201.

Wang, H. 1974. *From Mathematics to Philosophy*. New York: Humanities Press.

Wang, H. 1987. *Reflections on Kurt Gödel* . Cambridge, Mass.: MIT Press.

Weizenbaum, J. 1976. *Computer Power and Human Reason: From Judgment to Calculation*. San Francisco, Calif.: Freeman.