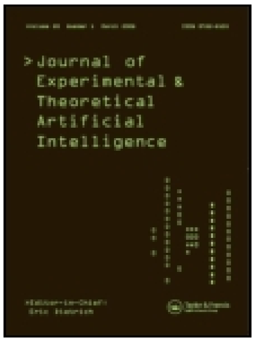

Journal of Experimental & Theoretical Artificial Intelligence

ISSN: 0952-813X (Print) 1362-3079 (Online) Journal homepage: http://www.tandfonline.com/loi/teta20

# Models, simulations, instantiations, and evidence: the case of digital evolution

Robert T. Pennock

*Journal of Experimental & Theoretical Artificial Intelligence*
Vol. 19, No. 1, March 2007, 29–42

# Models, simulations, instantiations, and evidence: the case of digital evolution

ROBERT T. PENNOCK*

Lyman Briggs School of Science and Departments of Philosophy and Computer Science, Michigan State University, East Lansing, MI 48825, USA

(*Received 2 October 2006; in final form 15 December 2006*)

What is the difference between a simulation of X and simply another instance of X? Is there a point at which the ''virtual reality'' of a model becomes the real thing? This paper examines these questions using cases taken from recent developments in evolutionary engineering and artificial life research. By implementing the Darwinian mechanism and setting it to work on a design problem, scientists and engineers find that evolution not only can improve prior designs, but also produce novel technological solutions. Artificial life systems Tierra and Avida which operate at a higher level of abstraction than evolutionary engineering applications. I analyze simulation as a rational concept ''S simulates R'' and argue that it always includes some relevant property P, of R, that is captured but that there is always also some other that it omits, and that pragmatic factors fix what counts as relevant. The border between a simulation and an instance or R can change depending upon the context. I show that in one sense, evo-technology and artificial life simulate organic evolution, but in another relevant sense they are instances of evolution itself. Biologists can use such systems to experimentally test evolutionary hypotheses such as those involving the evolution of complex features and altruism. This analysis suggests lines for future research on broader questions about models classification and confirmation.

*Keywords*: Modeling; Simulation; Instantiation; Digital evolution; Evidence

## 1. Introduction

Scientists have long investigated natural systems using laboratory models, but in recent years they have also begun to do so using computer simulations. Indeed, modelling on a computer has become so prevalent that it has become common for

*Email: pennock5@msu.edu


ISSN 0952–813X print/ISSN 1362–3079 online 
http://www.tandf.co.uk/journals
DOI: 10.1080/09528130601116113

people to speak of *simulation* as though the term referred exclusively to computer models. However, although computing technology has greatly extended their power, simulations are really a more general kind of scientific model; scientists studied phenomena using simulations well before the advent of computers. For instance, aeronautical engineers simulated flight using scale models fixed in wind tunnels to test possible jet wing shapes. And, although it has become common to do so, it is also improper to call a model a simulation simply because it is produced using a computer. For instance, while we can properly say that a computer can now simulate aerodynamic properties, allowing one to test possible wing shapes without a wind tunnel, it would not make sense to say that a computer is 'simulating' a mathematical calculation when it added two numbers together; rather, it is actually performing a calculation. These are simple cases, but distinguishing a simulation from an actualization is not always so obvious.

What is the difference between a simulation of something and simply another instance of it? Is there a point at which the 'virtual reality' of a simulation becomes the real thing? And how can investigating some phenomena by means of a simulation provide evidential warrant for scientific inferences? In this paper I will begin to examine these questions using several cases taken from recent developments in evolution engineering and artificial life (A-life) research. The specific issue I will consider is whether these systems are just simulations of evolution, or whether they are something more. I will conclude that in one sense they can be seen as simulations, but that in a more interesting sense they are actual instances of evolution, and that in both cases they can provide information about evolutionary processes.

I will not focus here on questions related to modelling the core evolutionary thesis of common descent. Neither will I discuss questions about the definition of biological life or the issue of whether the property of 'being alive' can be properly applied in the non-biological cases at hand, although the general questions in the paper could also be addressed through such an avenue. Rather, the common feature of the cases on which I will focus is what Darwin proposed as the *mechanism* that drives the evolution and adaptation of biological life.

Put briefly, the Darwinian mechanism works in the following manner. Organisms always vary from individual to individual in their specific properties, and new heritable variations randomly arise in each generation. In the constant struggle against limited resources in the environment and with each other, individuals whose distinctive variations give them a competitive edge over others are more likely than their rivals to survive and reproduce. Thus the properties that gave them that advantage will be naturally selected and passed on to their offspring. Darwin showed how iteration of these mechanisms produces species that are adapted to their environments.

Given that Darwinian processes gave rise to the complex functional adaptations we find in the biological world, we should expect that they could produce similarly useful results in other situations. This insight has recently begun to be used for various kinds of technology development. By implementing the Darwinian mechanism (typically, though not necessarily, in a computer), and setting it to work on a design problem, scientists and engineers are confirming that evolution not only can help to optimize prior designs, but can also produce novel design solutions. Although still in its infancy, evolutionary engineering is already being used in the design of products from computer circuits and software to jet engines and wings.

'Evo-technology', as some are beginning to call this form of engineering, is becoming a significant area of research with a wide range of practical applications. Similarly fruitful research is going on in the field of A-life, which is also based on insights from evolutionary theory. Indeed, biologists and other scientists are beginning to use A-life systems as models to perform experiments in order to study evolutionary principles (Lenski *et al.* 1999, Adami *et al.* 2000, Adami 2003, Wilke and Adami 2003, Goings *et al.* 2004).

Sometimes scientists do call these systems 'simulations' of evolution. However, others have made the stronger claim that they are not simulations but the real thing. Some go so far as to claim that we should recognize evolving digital organisms as a new life form. I will reserve my consideration of the question of what life is and whether it is reasonable to regard these as truly alive for a forthcoming occasion. Here I focus on the prior issue. How can we decide whether we should regard such processes as simulations or instances of evolution? To help answer that question, let us first consider some more general cases.

## 2. Simulation versus instance

There is a wide range of kinds of models that play a role in scientific reasoning, many, though not all, of which fall into the category of simulations. Abstract mathematical models of physical phenomena have been the most studied, but visual models such as diagrams and maps are equally if not more important. In experimental settings, scientists often pick one representative item upon which to focus research, as when biologists use *Drosophila* as a model organism for genetic investigations or frogs as models for teaching basic anatomical techniques and principles. There are also technological scale models, such as one finds in some of Leonardo da Vinci's ingenious inventions in the Museum of Science and Technology in Milan. Architectural scale models, such as that of Brunelleschi's dome, serve to display both engineering and aesthetic possibilities for evaluation. I follow Giere's *representational view* in holding that such scientific models are primarily, although not solely, 'tools for *representing the world*' (Giere 1999, p. 44), but I put special emphasis on their use as tools, and especially on the causal and pragmatic elements of that concept.

Like other kinds of modelling, simulation is a relational concept. Although we often speak of some 'simulation S' *simpliciter*, this is always properly shorthand for the more complete 'S simulates R,' where R is some real object or phenomenon (or kind of object or phenomenon) of interest. Studying R by means of a simulation S is similar to, but also clearly different from, studying it by means of direct instances of it. For example, we understand the difference between dissecting a real frog and using the Virtual Frog dissection simulation.

As a caveat, we should not assume that 'simulation' in scientific and engineering contexts has the same connotations as in ordinary speech, in particular, the trivializing connotation of being 'merely' a look-alike or 'just' an appearance. Computer simulations for some scientific purposes are already highly accurate and are becoming more so. In any case, these trivializing connotations probably became attached to the concept because, until recently, most simulations were indeed pale comparisons of the real thing, and it seems that that sense is already fading with the rapid development of virtual reality (VR) technologies. Computer-aided architectural design systems which combine three-dimensional renderings with VR display technology have developed to the degree that one can, virtually, move through proposed buildings to judge interior and exterior spatial proportions, lighting effects, and so on, turning one's head to see everything from the perspective one would have if one were examining a built structure. Nor are we any longer limited to the visual or aural; VR technologies for other sensory modalities, including the tactile, are also being developed.

I am focusing mostly upon engineering and technological examples here, but the points also hold for other kinds of scientific simulations. A simulation shares certain relevant causal properties with that which it simulates, but always lacks some other relevant causal relationships. What counts as relevant depends upon pragmatic factors, such as the question that one is asking or, more generally, the interests of the researcher constructing the simulation. The shared features might, for instance, be some structural or formal properties. These might be modelled in the same material as the original, or in quite a different substrate, as in simulating a wooden building by means of a physical scale model constructed of balsa, or by means of a computer rendering. In both cases, we capture certain causal features, particularly those associated with form, shape, and proportion, while setting aside other features such as the building's real size and mass. With different goals in mind, we might require those features as well, and perhaps more or fewer.

Thus, in the general case, to say that S simulates an R, we must have in mind some property P of R that we take S to capture, in contrast with some other property Q (or, usually, multiple properties) that S leaves out. Understanding this, we see why disputes about whether something is a good simulation or not often revolve as much around questions about what the relevant properties are, as much as whether they have been adequately captured. Similarly, if S shared *all*, or at least enough of, the properties that we take to define something as an R, then we see why S would not be a simulation, but just another instance of an R.

Things become more complicated if we try to move to second-order issues. When does it make sense to speak of a simulation of a simulation? What would a simulation of a map be, for example? A false map, i.e. one that contains errors—mislabelled streets, misplaced bridges, missing exits—is still a map, although we would say that it is a poor one. In some contexts, we might take a map that is intentionally deceptive, a fake map, still to be a map in the same way. However, with a different purpose in mind, we might indeed regard it as a simulation, in that its markings appear to, but do not in fact, map onto anything. Again, I would argue that careful analysis of pragmatic factors in particular settings can help to sort out such apparent difficulties.

In particular, it is usually the pragmatics of causal relations that is most significant in clarifying the difference between a simulation and an actualization of some phenomenon. We can see the elements of difference by comparing a real airplane flight with one in a flight simulator used to train pilots or a computer simulation to test wing design. In the flight simulator, one set of causal processes—the real relations that govern the interactions between pilot and plane in the air—have been replaced by a different set which shares enough structure to mimic the first in the manner of interest. For example, in both the real plane and the simulator, pulling back too hard on the stick has the effect of putting the plane in a stall, although in the latter it is, naturally, a simulated stall.

The same pattern holds when considering the use of models and simulators for airplane wing design. Aeronautical engineers traditionally designed wing shapes and tested them using small-scale models in a wind tunnel. As the physics of lift became better understood and became expressed mathematically, much of the testing shifted from physical models to computer simulations. Today, a proposed wing shape is geometrically represented in the computer and its flight characteristics are tested in a simulated environment which mathematically models the force of gravity, air resistance, and so on as the pilot takes it through its paces. Although, of course, there is neither gravity nor wind in the computer, the simulation captures the direction and strength of these forces.

Things become more interesting when we turn to evolutionary computer models and our original question about these, but before doing so we must briefly examine how scientific models, both actualizations and simulations, can be said to provide evidence about the empirical world.

## 3. Models and evidence

Scientific investigation and experimentation aim to increase our knowledge about the hidden features of the world, whether these be the aerodynamic laws that govern flight, or the biological laws that produce functional complexity. The evidence relation is supposed to say what data will tell us what we want to know but do not yet know, and, seen in this light, evidence may be thought of as information. A relevant bit of data is a piece of information. Confirmation theory typically discusses science simply in terms of the confirmation of hypotheses, but it may be more useful to think in broader terms; science is in the business of trying to find answers to empirical questions. Think of the workaday scientist in a period of normal science, as described by Kuhn. Such a scientist may be engaged in measuring an important constant and the question at hand is what its value is to within a certain degree of accuracy. Or the investigator may be faced with an anomaly in the current paradigm—a puzzle to be solved to which no hypothesis may immediately suggest itself, only questions such as 'What is the explanation of this unexpected phenomenon?' Altruistic behaviour, for instance, has seemed to be an anomaly for evolutionary theory. Often, of course, the scientist does have in mind a possible answer to a given question and proceeds by testing that hypothesis, but in many cases the scientist has advanced only to the point of having identified a problem and posed a question for investigation. Is there a black hole in the Markarian 335 galaxy? Why did the dinosaurs die out at the end of the Cretaceous period? What is the internal structure of the Earth? Questions are requests for information, and scientific questions such as these are no exception.

Once we think of science as posing empirical questions—requesting information from nature—we recognize that the observational and experimental tests that scientists use to answer those questions are attempts to extract information. We build the Hubble telescope to obtain better information about the swirling gases of Markarian 335, we make measurements of the distribution of iridium in rock layers at the K–T boundary to obtain information about events at end of the Cretaceous period, and we study the patterns of seismic waves to obtain information about the layers within the Earth. Scientists focus upon those data generated by processes that will be informative regarding the question at hand; in the cases above they do not bother looking at data on oil well drilling in Alaska, or critical reviews of the latest dinosaur movie, or Billboard's recent rock music charts, because such data are evidentially irrelevant. In such investigations we distinguish those data that count as evidentially relevant from those that do not on the basis of whether or not they are informative regarding the question at hand: a datum can be evidence of some state of affairs, A, because it can provide information about A. It is because evidence E contains information about some hypothesized empirical matter H that it can be cited in support of H.

Thus to say that we confirm (or gain empirical support for) a hypothesized structure in the world (say, an unobservable) by some evidential data (say, an observation) is to say that the evidence provides us with information about that structure. I have argued elsewhere that the confirmation relation, i.e. the relevance relation by virtue of which E is evidence of H, should be explicated in terms of ontic causal relations; the evidential inferences we may make are those that are licensed by the causal relations (Pennock 1991, 1995, 1998, 2004). This way of looking at evidential relevance helps make clear what is going on when scientists try to answer questions and test hypotheses using models. It also helps to show why simulations as well as instantiations can be of evidential value.

Evidential inferences drawn from simulations are analogical. Like all analogical arguments, their warrant depends upon the degree of similarity to that which they are modelling. The researcher must always carefully make a case for how a simulation S is similar to R with regard to the property P that is of interest. To the degree that a simulation is sufficiently similar to R, one may obtain useful information that helps one confirm hypotheses about R. However, in general, evidential inferences made from simulations are weaker than those drawn from instances because in the latter case the causal relationships of interest are directly realized.

Much more could be said about the evidence relation as it relates to these different kinds of models, but this is enough for our purposes in this paper, so we can now return to our case study.

## 4. Darwinian engineering

In evolutionary methods of design, engineers construct a computer model that can represent the problem space (e.g. wing shapes) and randomly mutate this to produce a large population of variants. They then allow a selection function to automatically weed out all but the most fit. The system keeps those variants that accomplish the set task somewhat better than others. Then it takes these winners, replicates them with new random mutations, and weeds them again, repeating this generation after generation. (One can also make the process more sophisticated by allowing variations to arise through random recombination, for example by randomly 'mating' winners.) Let me mention here just one example of a solution that was evolved using this sort of method.

At Stanford University, evolution of airplane wings in a computer came up with a novel shape. Imagine taking the wings of a plane and folding them about two-thirds of the way along their lengths so they point straight up, and then folding half or more of that length again so the tips point back horizontally toward the fuselage. There are several possible useful features of this design. For instance, a plane built with such wings would not need a tail. Moreover, its shorter wingspan would allow it to have a larger body that would carry 50% more passengers than a Boeing 747, while still meeting airport and manufacturing size constraints. (I give a more detailed explanation of Darwinian engineering in Pennock (1999, pp. 103–109).)

In the light of our previous discussion, we may now observe that while such systems only simulate wings and flying, they are not mere simulations of Darwinian evolution but are real instances of it. Although they differ in some of their particulars, they nevertheless instantiate the key elements of the Darwinian mechanism, given that one is asking certain kinds of questions, specifically ones about general evolutionary principles. The variation, reproduction, natural selection, and so on that produce evo-technology are, as Dennett (1995, p. 82) has put it, 'substrate neutral'. Let me now discuss this idea in greater detail in connection with two innovative A-life systems, Tierra and Avida, based on evolutionary computation.

## 5. Artificial life

The A-life environment Tierra was developed by ecologist Tom Ray to investigate the feasibility of an auto-adaptive genetic system (Ray 1991). Starting with a hand-coded organism whose program could mutate during replication, Ray observed the evolution of a small ecology of digital organisms, including parasites and hyperparasites. Inspired by Tierra's success, Avida is a second-generation system developed by Adami, Brown, and Ofria (Adami 1998) with greater flexibility to better serve as a platform for controlled experimentation on the artificial 'life forms' that inhabit its environment. Both Tierra and Avida operate at a higher level of abstraction than the evolutionary engineering applications we have discussed so far.

Simply put, in the virtual environment of computer memory, populations of digital organisms compete with each other for space and CPU time (which functions as their 'energy'). The systems are said to have a virtual 'chemistry' which is just their particular small set of component instructions. Avida has a basic set of about 24 instructions (depending upon the version) which are computationally (Turing) complete. An individual program is simply a string of these instructions (of no fixed length). Unlike some A-life systems where the goal may be to explicitly model, for example, fish or insects, the 'organisms' in Tierra and Avida are simply the computer programs themselves and nothing is done to make them represent particular biological species. However, the instruction set does allow them to exhibit a key property of biological beings, namely the ability to self-replicate. In both Tierra and Avida, an original self-replicating ancestor program is hand-coded and seeded in the environment, but thereafter it and its descendants are left to replicate and evolve on their own.

Evolution can happen to the digital organisms because errors are periodically allowed to occur in the programs. Random mutations (including substitutions, insertions, and deletions) of instructions may occur at any point in the string. If the variants that arise are still able to self-replicate, those changes will be inherited by their 'offspring'. Natural selection can occur in various ways, such as by having a 'reaper' automatically delete programs whose performance is poorer than others relative to some measure of fitness, such as efficiency of replication. Avida enriches possible fitness functions by including a single logical operator NAND (not–and) in its instruction set, but then allowing programs that acquire the ability to perform other more complex logical operations to be rewarded. Turned loose in their computational environments, over time these digital organisms evolve novel configurations and behaviours.

Of course, several of these elements are clearly simulated rather than actual, and we sometimes note that by putting terms in quotation marks. For example, we can speak of digital organisms as having a 'metabolism' in that they may need to process numbers logically to acquire additional 'energy', but these concepts are not instantiated in the model in the way that they are in real life, but are represented by means of analogous processes. The 'energy' that these digital organisms use is actually central processing unit cycles, i.e. processing time to execute instructions, which is divided up by a time-slicing mechanism so that the organisms can operate essentially in parallel. This functions as a good analogue, but we would not want to say that they are instances of real energy uptake by a biological metabolism.

Some concepts could be thought of as simulations or as actualizations, depending upon what one takes to be the character of what is being modelled. For instance, if one focuses on the physical components of biological organisms and takes the term *organism* to refer essentially to their organic chemistry, one should speak of digital 'organisms' as simulations because they are not composed of the relevant biomolecules. However, if one thinks instead of biological organisms at a slightly higher degree of abstraction and focuses upon their characteristic functional organization (as is reasonable given that we consider it possible that we might one day discover biological organisms on other planets that have a different chemistry to those on Earth), then one can drop the quotation marks because digital organisms can be seen to be instances of the same sort of organized functional complexity.

Returning to our original question, look now at what is going on with respect to the basic Darwinian processes. Recall that these are variation, inheritance, and natural selection. Time is usually taken for granted, but it is worth mentioning explicitly here. I will discuss these briefly in turn.

(i) **Variation**. The digital organisms are simulating the variation of organic organisms; we do not see varieties of fur, feathers, skin, and scales, or even varieties of proteins or biomolecules. However, they are not simulating variation itself—they exhibit real differences. For instance, in an evolving population of Tierrans or Avidians, one may observe phenotypic variants such as differential lengths of time per generation or differential ability to process or 'metabolize' resources. An evolutionary run in Avida typically begins with an organism that can perform no metabolic tasks, but as the run proceeds one soon sees the evolution of organisms that can perform various tasks. One can also observe the genotypic variation that underlies these phenotypic differences, and see how random mutations produce changes in the genetic material. As noted, genotypic variation may arise from point mutations, insertions, deletions, and, in sexual populations, recombination.

(ii) **Inheritance**. Although digital organisms do not reproduce using the types of biochemical processes that organic beings use, they do actually reproduce themselves. Likewise, the resulting inheritance of characters from one generation to the next is no simulation—these characters are truly passed on. In the digital environment of a computer core, reproduction is essentially a process of copying into a new memory location. For a digital organism to successfully self-replicate it must measure itself, allocate an appropriate space in memory, and copy its genome instruction by instruction, including the sequence of instructions that governs this whole process. Inheritance of genotypic (and thus phenotypic) characteristics occurs to the degree that the copying process is faithful. Of course, because random mutations in the instructions may occur during the replication process, inheritance is imperfect as it is in natural systems. Indeed, it often happens that mutations will change instructions that were used in the replication process and so result in an organism that is not viable. However, occasionally mutations occur that improve the efficiency of reproduction and then this improvement can be inherited by subsequent generations.

(iii) **Natural selection**. Similarly, one may observe the same struggle to survive as is seen in biological organisms; this is an actual competition and no mere mimic of natural selection. As noted above, variants can arise that can perform tasks that their rivals cannot, or ones that perform the same task but more efficiently than another. Other things being equal, a mutational variant that happens to replicate faster will out-compete others in the population and so will tend to predominate in subsequent generations. In Avida, an organism that evolves the capacity to perform tasks that give it additional processing time will similarly be naturally selected over less fit organisms in the population.

Some might say that we should put ‘survive’ in quotation marks above, arguing that the notion of survival is essentially tied to the concepts of life and death, and that digital organisms cannot survive if they can neither live nor die in the biological sense. Again, I am bracketing the question of whether (and which) digital organisms might be said to be alive in a biological sense. Researchers who argue the strong A-life position would say that there are already digital organisms that qualify as being alive. But even without taking a stand on that question, it is reasonable to decouple the notion of survival from its connection to a literal notion of life and death, and to use the term in its more abstract sense as we do in speaking of ideas that have survived the test of time. We commonly use the term in a biological sense to speak of traits that survive in a population, even when the specific traits one is referring to are not themselves alive. In this sense of the term, digital organisms (and their characteristics) may be said to differentially survive and reproduce as they are more or less adapted to their environment. However, as seen above, whether or not we use this way of expressing the idea, it is clear that natural selection can be perfectly instantiated in A-life systems. Again, this is not a simulation.

(iv) **Time**. In the real world organisms compete against one another and are naturally selected in real time. What about time in a computer? This is an interesting technical and conceptual question but the details do not concern us here. In a serial computer, of course, processing occurs sequentially and so an A-life environment must be set up with a 'time-slicing' process so that organisms may advance in parallel. One might argue that for this reason, strictly speaking, time in its general sense is only simulated but not actualized. But, again, one must look carefully at the pragmatics of a question to differentiate simulation from an actualization. For certain kinds of questions it would indeed be important to recognize this internal time as a simulation of real time, but our interest here is the question of whether A-life systems instantiate the Darwinian mechanism. The relevant sense of time for an evolutionary system is not the passage of clock time *per se*, but rather the passage of generational time—the iterated process of variation, selection, and reproduction in parallel—and this is properly realized in both Tierra and Avida.

Thus, in the relevant Darwinian sense, in studying the descent with modification of A-life in Tierra and Avida, scientists are observing evolution itself.

Some interesting differences remain between these A-life worlds and natural biological systems. For example, at least in the current versions of Tierra and Avida, the digital organisms are what we might call 'bare genomes'—their genetic instructions interact directly with the environment and each other, without being clothed in a distinct 'body'. Although this models something rather like the situation of what origin-of-life researchers hypothesize may have been a primitive RNA-world, it omits an important feature of biological organisms; we should not expect the systems to exhibit those kinds of complex properties that arise from the causal relationships between a genome and the body that carries it. Remember, however, that Tierra and Avida nevertheless both do model the distinction between genotype and phenotype, with natural selection acting by virtue of the latter, rather than directly on the former. In Avida, for instance, differential selection can depend upon the phenotypic property of what or how many logical operations a digital organism performs. The omission could easily be remedied if, for example, the programs were embedded in the appropriate way in physical or virtual robots. More speculatively, we might be able to find a sense in which Avida's evolved logical or other operations could be said to function as a kind of virtual body. I will not pursue such interesting questions here, but simply call attention to the fruitfulness of thinking along such lines about the possible boundaries between a simulation and what is simulated. In such investigations, scientists may test the utility and scope of proposed theoretical concepts and principles.

## 6. Objections considered

There are several objections one might make to the conclusion that in studying evo-technology and the A-life in Tierra and Avida, biologists are observing evolution itself. Here I just have space to briefly dismiss one naive objection and then deal in rather more detail with a second.

One objection, which is regularly made by anti-evolutionists who argue that only intelligent design can produce novel complex adaptations, is that A-life and similar systems are not truly Darwinian systems because they are made by programmers who are intelligent agents and that whatever complex features supposedly evolved were really already put there by their intelligent action. This kind of objection exhibits basic misunderstandings about how such systems work. I have shown elsewhere (Pennock 2000) how the novel results that evolutionary systems can produce are not in any relevant way 'designed in' by the programmers, and so I will not repeat the explanation here.

A second objection is that neither evo-technology nor A-life qualifies as evolution because evolution is organic, while they are not. Genetic evolution depends on DNA and RNA, and if we sometimes speak of non-biological systems as evolving, that can only be in a loose metaphorical sense. Thus these systems are simulations, not instances, of evolution.

I want to argue that this is the wrong way to think about the situation. The objection betrays an unwarranted prejudice in favour of the organic in the traditional sense of that term. While it is true that evolution of carbon-based organisms is the prototype of the concept, this historical fact is not a sufficient reason to limit its scope. Why be carbon-centric? It is the patterns of causal interactions that are relevant, not the particular material substrate. It does not make sense to tie evolution to a carbon substrate in the same way that that it would not make sense to think that wings must be composed of bones, flesh, and feathers. For our aeronautical interests, it is primarily the shape that counts in making something an instance of a wing, and with the advancement of technology we properly recognized as wings those constructed of balsa wood, plastic, aluminum, and so on. Similarly, the material substrate of the Darwinian processes should be irrelevant to whether we recognize something as an instance of Darwin's evolutionary mechanism.

We may also give some historical reasons to question the premise of the objection. While we now know that DNA and RNA are the substrate for genetic evolution, Darwin of course did not. His account of the mechanism of descent with modification was described, by necessity, at a higher level of generality. Darwin was dealing just with organic beings in the traditional sense, but the causal processes he identified as the driving forces of evolution made no essential reference to that. He spoke of organisms that randomly varied, of inheritance, and of a struggle for existence that led to natural selection, but nothing about those processes require that the organisms be organic. Even in the modern synthesis, when Darwinism was connected to Mendelism, there was no explicit mention of what the genetic material was; that knowledge would not be discovered until years later. Mendelian genetics spoke only of an abstract notion of genes that obeyed generic laws of segregation and assortment. If in a nearby possible world it had turned out that Watson and Crick discovered that genetic information was coded in a sulphur- or copper-based molecule rather than in DNA, that would not have affected the Darwinian account in the least.

## 7. Evidence from A-life

We are now in a position to see how an evolving A-life environment like Avida can be of value as an experimental system for answering certain sorts of biological questions and for testing evolutionary hypotheses. I will briefly describe two different studies, one dealing with the evolution of complex traits and the other with the evolution of altruism.

In the *Origin of Species*, Darwin (1859) not only presented the positive evidence for descent with modification, but also tackled what appeared to be the most significant objections to the theory. Among these was the problem of the origin of new features, especially those that he called 'organs of extreme perfection and complication' (Darwin 1859, p. 186). The key issue was the complexity of these organs. How could complex traits such as the mammalian eye arise? Darwin's answer to this problem had three main components. He granted that it was unlikely for such complex structures to appear all at once, and argued that they must evolve gradually through a series of intermediate stages. Secondly, he proposed that these intermediates would themselves be selectively advantageous, but not necessarily in the same way—functions would probably convert from one to another. Thirdly, he argued that a generic complex feature, such as the eye, could evolve via many alternative pathways. Of course, the imperfections of the fossil record made it unlikely that one would be able to reconstruct the evolution of a particular complex feature. Nor are direct demonstrations in natural systems possible, because the processes are too slow relative to the lifetime of a scientist. Thus Darwin and scientists afterwards had to provide indirect evidence by taking comparative data from various species of the same group, and although there was ample evidence of this sort, the ideal would be to have, in addition, an experimental test that looked directly at lineal ancestors. A-life models made such a test possible.

To directly test Darwin's proposed solution to the objection to complex features, we set up a controlled experiment in which one set of populations evolved in an environment that rewarded only one very complex function—in particular, the logical operation *equals*—and another set evolved in environment that also rewarded simpler functions (Lenski *et al.* 2003). In the latter conditions we observed populations that did evolve the *equals* trait. With Avida, however, there was no problem of an incomplete fossil record; we could keep a complete trace of how the trait evolved along a line of descent, seeing how natural selection acted upon random mutations in the genomes of organisms in the population. We observed all the patterns that Darwin had predicted, confirming his explanations in a more fine-grained manner than had previously been possible. More than that, we also discovered an additional significant pattern: the sequence of steps that resulted in the trait did not necessarily improve an organism's fitness at each step; in some cases mutational steps were temporarily deleterious, but wound up fortuitously setting up a subsequent beneficial step. Thus the digital experiment was not only able to test Darwin's hypotheses directly, but also allowed a novel discovery.

The A-life model allowed this kind of experimental test because it exactly instantiated the elements of the Darwinian mechanism. Nothing in Darwin's explanations made any reference to any particular biochemical substrate; the relevant explanatory level involved generalized notions of functions arising by means of the general cause-and-effect processes of Darwin's law of evolution by natural selection.

In the same way, we are now able to use Avida to test hypotheses about the evolution of altruistic behaviour. As mentioned above, altruism was taken to be a puzzle for evolutionary theory. How could what seems to be an inherently 'selfish' genetic process produce selfless behaviour? William Hamilton provided the basis for a solution to this puzzle in his theory of *inclusive fitness* which showed how natural selection can favour an altruist gene when the cost $C$ to the altruist is less than the benefit $B$ conferred on the recipient multiplied by the degree of relatedness $R$ between the individuals (Hamilton 1964). Hamilton's notion is now commonly referred to as *kin selection* and it is easiest to understand in this kind of context how altruistic behaviour towards kin, who are likely to share one's altruist gene, can indirectly cause the gene to be selectively favoured, even though it is passed on through a relative rather than directly by oneself. Extensions of this idea have posited that more accurate ways of determining the presence of the altruist gene, perhaps by means of a linked physical sign—for instance, a green beard (Dawkins 1976)—could theoretically out-compete kin-based altruism.

To test these and our own ideas about the evolution of altruism, we added some special instructions to the basic Avida instruction set that would allow organisms to donate energy to others under different kinds of conditions, such as degree of kinship, degree of genetic similarity, the presence of a donated gene in a possible recipient, and so on. Again, the relevant level of generality of these concepts of inclusive fitness, kin selection, and so on makes no reference to any particular instantiation of the evolutionary mechanism. For instance, Hamilton's relatedness variable $R$ is a theoretical concept that is as equally actualized in Avida as in natural biological systems. The specific implementation of the sensor by which an organism identifies which organisms are kin (or relevantly similar in some other way) is simply a simulation of the various processes that may be found in nature and is not meant to instantiate any particular sensory system. However, this sort of simulation is sufficient to license evidential inferences because it sufficiently captures the relevant functional property. As before, experiments with Avida were able to both test and extend existing biological theory.

## 8. Conclusions

In this paper, I have considered the nature of simulation models in general, and the status of evo-technology and A-life systems in particular. I began with the idea that simulation is a relational concept and that, like other representational models in science, simulations are tools. I argued that the distinction between a simulation of R and an instance of R depends upon what relevant causal relationships are present or absent, and that what counts as relevant depends upon which pragmatic values govern in a given context. For some phenomena, for some purposes, the distinction between virtual and real all but vanishes. For general questions about the power of the Darwinian mechanism, evolutionary computations systems are true experimental systems and should not be thought of simply as simulations; they are actual instances of the processes of evolution and license evidential inferences of the same warrant as natural systems.

This paper represents one element of a larger project that seeks to investigate the evolution of intelligent behaviour as an alternative to traditional top-down AI approaches. In using evolutionary computation to model this, we must be clear about the ways in which evolution may be recognized as a general abstract theoretical structure that is not limited to biology. What we have seen shows how we may make progress in that analysis, and begin to determine just how far evolutionary processes extend beyond the organic. Future work must examine the viability of these basic concepts in greater detail and in other contexts, and extend the analysis to other related evolutionary concepts. This paper is a part of a larger investigation of the nature of evidence and general issues about models and confirmation, such as what types of relations between model and modelled allow the former to serve as a base for inferences about the latter. In scientific contexts, whether theoretical or applied, our core interests do primarily involve such questions about what are warranted empirical inferences. For a model to serve as a basis for evidential inferences, it must adequately capture those features of the real system that are under investigation. Thus how we classify features of the world and how we use evidence to confirm generalizations regarding them are closely interlinked. These ideas and their relationships will have to wait for further elaboration at a later time.

## References


C. Adami, *Introduction to Artificial Life*, New York: Telos, 1998.

C. Adami, "Sequence complexity in Darwinian evolution", *Complexity*, 8, pp. 49–56, 2003.

C. Adami, C. Ofria and T.C. Collier, "Evolution of biological complexity", *Proc. Natl. Acad. Sci. USA*, 97, pp. 4463–4468, 2000.

C. Darwin, *On the Origin of Species by Means of Natural Selection*, London: Murray, 1859.

R. Dawkins, *The Selfish Gene*, Oxford: Oxford University Press, 1976.

D.C. Dennett, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, New York: Simon & Schuster, 1995.

R.N. Giere, "Using models to represent reality", in *Model-Based Reasoning in Scientific Discovery*, L. Magnani, N.J. Nersessian and P. Thagard, Eds, New York: Kluwer Academic Plenum Press, 1999, pp. 41–57.

S. Goings, J. Clune, C. Ofria and R. Pennock, "Kin-selection: the rise and fall of kin-cheaters", in *Artificial Life IX: Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems*, 2004.

W. Hamilton, "The genetical evolution of social behavior", *J. Theor. Biol.*, 7, pp. 1–31, 1964.

R.E. Lenski, C. Ofria, C. Adami and T.C. Collier, "Genome complexity, robustness and genetic interactions in digital organisms", *Nature*, 400, pp. 661–664, 1999.

R.E. Lenski, C. Ofria, R.T. Pennock and C. Adami, "The evolutionary origin of complex features", *Nature*, 423, pp. 139–144, 2003.

R.T. Pennock, *Causally Licensed Inference and the Confirmation Relation*, Pittsburgh: University of Pittsburgh, 1991.

R.T. Pennock, "Epistemic and ontic theories of explanation and confirmation", *Philos. Sci. (Jpn)*, 28, pp. 31–45, 1995.

R.T. Pennock, "Evidential relevance and the Grue Paradox", *Philos. Sci. (Jpn)*, 31, pp. 101–119, 1998.

R.T. Pennock, *Tower of Babel: The Evidence Against the New Creationism*, Cambridge, MA: MIT Press, 1999.

R.T. Pennock, "Can Darwinian mechanisms make novel discoveries?: learning from discoveries made by evolving neural networks", *Found. Sci.*, 5, pp. 225–238, 2000.

R.T. Pennock, "Bayesianism, ravens and evidential relevance", *Ann. Jpn Assoc. Philos. Sci.*, 13, pp. 1–17, 2004.

T.S. Ray, "An approach to the synthesis of life", in *Artificial Life II*, C. Langton, C. Taylor, J.D. Farmer and S. Rasmussen, Eds, Santa Fe, NM: Addison-Wesley, 1991, pp. 371–408.

C.O. Wilke and C. Adami, "Evolution of mutational robustness", *Mutat. Res.*, 522, pp. 3–11, 2003.