# Katherine Hayles: Narratives of Artificial Life

 In contrast to the circular processes of Maturana's autopoiesis, the figure most apt to describe the third wave is a spiral. Whereas the second wave is characterized by an attempt to include the observer in an account of the system's functioning, in the third wave the emphasis falls on getting the system to evolve in new directions. Self-organization is no longer enough. The third wave wants to impart an upward tension to the recursive loops of self-organizing processes so that, like a spring compressed and suddenly released, they break out of the pattern of circular self-organization and leap outward into the new.

 Just as von Foerster served as a transition figure between the first and second waves, so Francisco Varela bridges the transition between the second and third waves. We saw in Chapter 6 that Maturana and Varela extended the definition of the living to include artificial systems. After co-authoring *The Embodied Mind*,[1] Varela began to work in a new field known as Artificial Life and co-edited the papers from the first European conference in that field. In the introduction to the conference volume, *Towards a Practice of Autonomous Systems*, he and his co-author Paul Bourgine lay out their view of what Artificial Life should be.[2] They locate its origins in cybernetics, referencing Grey Walter's electronic tortoise and Ross Ashby's homeostat. Although some characteristics of autopoiesis are reinscribed on the successor field of Artificial Life, especially the idea that systems are operationally closed, other features are new. The change is signaled in Varela's subtle reconception of autonomy. He and his co-author write, "Autonomy in this context refers to [the living's] basic and fundamental capacity to <u>be</u>, to assert their existence and bring forth a world that is significant and pertinent without being pre-digested in advance. Thus the autonomy of the living is understood here both in regards to its actions and to the way it shapes a world into significance. This conceptual exploration goes hand in hand with the design and construction of autonomous agents and suggest an enormous range of applications at all scales, from cells to societies" (p. xi). For Maturana, "shap[ing] a world into significance" meant that perception was linked primarily to internal processes rather than external stimuli.[3] We have seen the difficulties he had with evolution, because he sought to put the emphasis instead on the organism's holistic nature and autopoietic circularity. When Varela and his co-author speak of "shap[ing] a world into significance," the important point for them is that the system's organization, far from remaining unchanged, can transform itself through emergent behavior. The change is not so much an absolute break, however, as a shift in emphasis and a corresponding transformation in the kind of questions the research programs pose, as well as new strategies for answering them. Thus the relation of the third wave to the second is again one of seriation, an overlapping pattern of replication and innovation.

The shift in questions and methodologies is not, of course, neutral. For researchers who come to the field from backgrounds in cognitive science and computer science, rather than from autopoiesis as Varela does, the underlying assumptions all too easily lend themselves to reinscribing a disembodied view of information. But as Varela's presence in the field indicates, not everyone who works in the field agrees that disembodied "organisms" are the best way to construct Artificial Life. Just as there were competing camps in the Macy conferences, one arguing for a disembodied view of information and one for a contextualized embodied view, so in Artificial Life some researchers concentrate on simulations, insisting that embodiment is not necessary, whereas others argue that only embodied forms can fully capture the richness of an organism's interactions with the environment. Our old friend the observer, who was at the center of the epistemological revolution Maturana sparked, in the third wave retreats to the periphery, with a consequent loss of the sophistication that Maturana brought to epistemological questions. The observer has, however, not altogether vanished from the scene, remaining in the picture as narrator and narratee of stories about Artificial Life. To see how the observer's presence helps to construct the field, let us turn now the consider the strange flora and fauna of the world of Artificial Life.

## **The Nature and Artifice of Artificial Life**

At the Fourth Conference on Artificial Life in the summer of 1994, evolutionary biologist Thomas S. Ray put forth two proposals.[4] The first was a plan to preserve biodiversity in Costa Rican rain forests; the second was a suggestion that Tierra, his software program creating Artificial Life forms inside a computer, be released on the Internet so that it could "breed" diverse species on computers all over the world. Ray saw the two proposals as complementary. The first aimed to extend biological diversity for protein-based life forms; the second sought the same for silicon-based life forms. Their juxtaposition dramatically illustrates the reconstruction of nature going on in the field of Artificial Life, affectionately known by its practitioners as AL. "The object of an AL instantiation," Ray wrote recently, "is to introduce the *natural* form and process of life into an *artificial* medium" (emphasis added).[5] The lines startle. In Ray's rhetoric, the computer codes comprising these "creatures" become natural forms of life; only the medium is artificial.

How is it possible in the late twentieth century to believe, or at least claim to believe, that computer codes are alive? And not only alive, but natural? The question is difficult to answer directly, for it involves assumptions that are not explicitly articulated. Moreover, these presuppositions do not stand by themselves but move in dynamic interplay with other formulations and ideas circulating through the culture. In view of this complexity, the subject is perhaps best approached through indirection, by looking not only at the scientific content of the programs but also at the stories told about and through them. These stories, I will argue, constitute a multilayered system of metaphoric and material relays through which "life," "nature," and the "human" are being redefined.

The first level of narrative with which I will be concerned is the Tierra program and various representations of it by Ray and others. In these representations, authorial intention, anthropomorphic interpretation, and the program's operations are so interwoven that it is impossible to separate them. As a result, the program operates as much within the imagination as it does within the computer. The second level of narrative focuses on the arguments and rhetorical strategies that AL practitioners use as they seek to position Artificial Life as a valid area of research within theoretical biology. This involves telling a story about the state of the field and the contributions that AL can make to it. As we shall see, the second-level story quickly moves beyond purely professional considerations, evoking a larger narrative about the kinds of life that have emerged, and are emerging, on earth. The narrative about the present and future of terrestrial evolution comprises the third level. It is constituted through speculations on the relation of human beings to their silicon cousins, the "creatures" who live inside the computer. Here, at the third level, the implication of the observer in the construction of all three narratives becomes explicit. To interrogate how this complex narrative field is initiated, developed, and interpolated with other cultural narratives, let us begin at the first level, with an explanation of the Tierra program.

Conventionally, Artificial Life is divided into three research fronts. *Wetware* is the attempt to create artificial biological life through such techniques as building components of unicellular organisms in test tubes. *Hardware* is the construction of robots and other embodied life forms. *Software* is the creation of computer programs instantiating emergent or evolutionary processes. Although each of these areas has its distinctive emphases and research agendas, they share the sense of building life from the "bottom up." In the software branch, with which I am concerned here, the idea is to begin with a few simple local rules, then through structures that are highly recursive, allow complexity to emerge

spontaneously. Emergence implies that properties or programs appear on their own, often developing in ways not anticipated by the person who created the simulation. Structures that lead to emergence typically involve complex feedback loops in which the outputs of a system are repeatedly fed back in as input. As the recursive looping continues, small deviations can quickly become magnified, leading to the complex interactions and unpredictable evolutions associated with emergence.[6]

Even granting emergence, it is still a long jump from programs that replicate inside a computer to living organisms. This gap is bridged largely through narratives about the programs that map them into evolutionary scenarios traditionally associated with the behavior of living creatures. The narratives translate the operations of computer codes into biological analogues that make sense of the program's logic. In the process, the narratives transform the binary operations that, on a physical level, amount to changing magnetic polarities into the high drama of a Darwinian struggle for survival and reproduction. To see this transformation in action, consider the following account of the Tierra program. This account is compiled from Thomas Ray's published articles and unpublished working papers, conversations I had with him about his program, and public lectures he has given on the topic.[7]

When I visited him at the Santa Fe Institute, he talked about the genesis of Tierra. Frustrated with the slow pace of natural evolution, he wondered if it would be possible to speed things up by creating evolvable artificial organisms within the computer. One of the first challenges he faced was designing programs robust enough to withstand mutation without crashing. To induce robustness, he conceived of building inside the regular computer a "virtual computer" out of software. Whereas the regular computer uses memory addresses to find data and execute instructions, the virtual computer uses a technique Ray calls "address by template." Taking its cue from the topological matching of DNA bases, in which one base finds its appropriate partner by diffusing through the medium until it locates another base with a surface it can fit into like a key into a lock, address by template matches one code segment to another by looking for its binary inverse. For example, if an instruction is written in binary code 1001, the virtual computer searches nearby memory to find a matching segment with the code 0110. The strategy has the advantage of creating a container for the organisms that renders them incapable of replicating outside the virtual computer, for the address by template operation can occur only within a virtual computer. Presented with a string such as 0110, the regular computer would read it as data rather than instructions to replicate.

Species diversify and evolve through mutation. To introduce mutation, Ray creates the equivalent of cosmic rays by having the program flip a bit's polarity once in every 10,000 executed instructions. In addition, replication errors occur about once in every 1,000 to 2,500 instructions copied, introducing another source of mutation. Other differences spring from an effect Ray calls "sloppy reproduction," analogous to the genetic mixing that occurs when a bacterium absorbs fragments of a dead organism nearby. To control the number of organisms, Ray introduced a program that he calls the "reaper." The "reaper" monitors the population and eliminates the oldest creatures and those who are "defective," that is, those who most frequently have made errors in executing their programs. If a creature finds a way to replicate more efficiently, it is rewarded by moving down in the reaper's queue and so becomes "younger."

The virtual computer starts the evolutionary process by allocating a block of memory that Ray calls the "soup," by analogy with the primeval soup at the beginning of life on earth. Inside the soup are unleashed self-replicating programs, normally starting with a single 80-byte creature called the "ancestor." The ancestor is comprised of three segments. The first segment counts its instructions to see how long it is (this procedure ensures that the length can change without throwing off the reproductive process); the second segment reserves that much space in nearby memory, putting a protective membrane around it (by analogy with the membranes enclosing living organisms); and the third segment copies its code into the reserved space, thus completing the reproduction and creating a "daughter cell" from the "mother cell." To see how mutation leads to new species, consider that a bit flip occurs in the last line of the first segment, changing 1100 to 1110. Normally the program would find the second segment by searching for its first line, encoded 0011. Now, however, the program searches until it finds a segment starting with 0001. Thus it goes not to its own second segment but to another string of code in nearby memory. Many mutations are not viable and do not lead to reproduction. Occasionally, however, the program finds a segment starting with 0001 which will allow it to reproduce. Then a new species is created, as this organism begins producing offspring.

When Ray set his program running overnight, he thought he would be lucky to get a one- or two- byte variation from the 80-byte ancestor. Checking it out the next morning, he found that an entire ecology had evolved, including one 22-byte organism. Among the mutants were parasites that had lost their own copying instructions but had developed the ability to invade a host and hijack its copying procedure. One 45-byte parasite had evolved into a benign relationship with the ancestor; others were destructive, crowding out the ancestor with their own offspring. Later runs of the program saw the development of hyperparasites, which had evolved ways to compete for time as well as memory. Computer time is doled out equally to each organism by a "slicer" that determines when it can execute its program. Hyperparasites wait for parasites to invade them. Then, when the parasite attempts to reproduce using the hyperparasite's own copy procedure, the hyperparasite directs the program to its own third segment instead of returning it to the parasite's ending segment. Thus the hyperparasite's code is copied on the parasite's time. In this way the hyperparasite greatly multiplies the time it has for reproduction, for in effect it appropriates the parasite's time for its own.

This, then, is the first-level narrative about the program. It appears with minor variations in Ray's articles and lectures. It is also told in the Santa Fe Institute videotape "Simple Rules . . . Complex Behavior," in which Ray collaborated with a graphic artist to create a visual representation of Tierra, accompanied by his voiceover.[8] If we ask how this narrative is constituted, we can see that statements about the program's operation and interpretations of its meaning are in continuous interplay with each other. Consider the analogies implicit in such terms as "mother cell," "daughter cell," "ancestor," "parasite" and "hyperparasite." The terms do more than set up parallels with living systems; they also reveal Ray's intention in creating an appropriate environment in which the dynamic emergence of evolutionary processes could take off. In this respect Ray's rhetoric is quite different from that of Richard Dawkins in *The Selfish Gene*, a work also deeply informed by anthropomorphic constructions.[9] Dawkins's rhetoric attributes to genes human agency and intention, creating a narrative of human-like struggle for lineage. In this construction, Dawkins overlays onto the genes strategies, emotions, and outcomes that properly belong to the human domain. Ray, by contrast, is working with artificial systems designed by humans *precisely so they would be able to manifest these qualities*. This is the primary reason explanation and interpretation are inextricably entwined in the first-level narrative. Ray's biomorphic namings and interpretations function not so much as an overlay as an explication of an intention that was there at the beginning. Analogy is not incidental or belated but central to the program's artifactual design.

Important as analogy is, it is not the whole story. The narrative's compelling effect comes not only from analogical naming but also from images. In rhetorical analysis, of course, "image" can mean either an actual picture or a verbal formulation capable of evoking a mental picture. Whether an image is a visualization or visually evocative language, it is a powerful mode of communication

because it draws upon the high density of information that images convey. Visualization and visually evocative language collaborate in the videotape the Santa Fe Institute made to publicize its work, entitled "Simple Rules . . . Complex Behavior. As the narrative about Tierra begins, the camera flies over a scene representing the inside of a computer. This stylized landscape is dominated by a block-like structure representing the CPU (Central Processing Unit) and dotted with smaller upright rectangles representing other integrated circuits. Then the camera zooms into the CPU, where we see a grid upon which the "creatures" appear and being to reproduce. They are imaged as solid polygons strung together to form three sections, representing the three segments of code. Let us linger at this scene and consider how it has been constructed. The pastoral landscape upon which the creatures are visualized instantiates a transformation characteristic of the new information technologies and the narratives that surround them. A material object (the computer) has been translated into the functions it performs (the programs it executes) which in turn have been represented in visual codes familiar to the viewer (the bodies of the "creatures"). The path can be represented schematically as material base--> functionality--> representational code. This kind of transformation is extremely widespread, appearing in popular venues as well as scientific applications. It is used by William Gibson in *Neuromancer,* for example, when he represents the data arrays of a global informational network as solid polygons in three-dimensional space that his protagonist, transformed into a point of view or pov, can navigate as though he were flying through the atmosphere.[10] The schematic operates in remarkably similar fashion in the video, where we become a disembodied pov flying over the lifeworld of the "creatures," comfortingly familiar in its three-dimensional spaces and rules of operation. Whereas the CPU landscape corresponds to the computer's interior architecture, however, the lifeworld of the creatures does not. The seamless transition between the two elides the difference between the *material* space inside the computer and the *imagined* space that, in actuality, consists of computer addresses and magnetic polarities on the computer disk.

To explore how these images work to encode assumptions, consider the bodies of the "creatures," which resemble stylized ants. In the program, the "creatures" have bodies only in a metaphoric sense, as Ray recognizes when he talks about their bodies of information (itself an analogy).[11] These bodies of information are not, as the expression might be taken to imply, phenotypic expressions of informational codes. Rather, the "creatures" *are* their codes. For them, genotype and phenotype amount to the same thing; the organism is the code, and the code is the organism. By representing them as phenotypes, visually by giving them three-dimensional bodies and verbally by calling them ancestors, parasites, and such, Ray elides the difference between behavior, properly restricted to an organism, and executing a code, applicable to the informational domain. In the process, assumptions we have about behavior, in particular thinking of it as independent action undertaken by purposive agents, are transported into the narrative.

Further encoding takes place in the plot. Narrative tells a story, and intrinsic to story is chronology, intention, and causality. In Tierra, the narrative is constituted through the story that emerges of the creatures' struggle for survival and reproduction. More than an analogy or an image, this is a drama that, if presented in a different medium, one would not hesitate to identify as an epic. Like an epic, it portrays life on a grand scale, depicting the rise and fall of races, some doomed and some triumphant, recording the strategies they invent as they play for the high stakes of establishing a lineage. The epic nature of the narrative is even more explicit in Ray's plans to develop a global ecology for Tierra. In his proposal to create a digital "biodiversity reserve," the idea is to release the Tierra program on the Internet so that it can run in background on computers across the globe. Each site will develop its own microecology. Because background programs run when demands on the computer are at a minimum, the programs will normally be executed late at night, when most users are in bed. Humans are active while the "creatures" are dormant; they evolve while we sleep. Ray points out that someone monitoring activity in Tierra programs would therefore see it as a moving wave that follows darkfall around the world. Linking the creatures' evolution to the human world in a complementary diurnal rhythm, the proposal edges toward a larger narrative level that interpolates their story into ours, our into theirs.

A similar interpolation occurs in the video. The narrative appears to be following the script of Genesis, from the lightning that flickers over the landscape, representing the life force, to the "creatures" who, like their human counterparts, follow the Biblical imperative to be fruitful and multiply. When a death's head appears on the scene, representing the reaper program, we understand that this pastoral existence will not last for long. The idyll is punctured by competition between species, strategies of subversion and co-optation, and exploitation of one group by another--in short, all the trappings of rampant capitalism. To measure how much this narrative accomplishes, it is helpful to remember that what one actually sees as the output of the Tierra program is a spectrum of bar graphs tracking the numbers of programs of given byte lengths as a function of time. The strategies emerge when human interpreters scrutinize the binary

codes that constitute the "creatures" to find out how they have changed and determine how they work.

No one knows this better, of course, than Ray and other researchers in the field. The video, as they would no doubt want to remind us, is merely an artist's visualization that has no scientific standing. It is, moreover, intended for a wide audience, not all of whom are presumed to be scientists. This fact in itself is interesting, for the tape as a whole is an unabashed promotion of the Santa Fe Institute. It speaks to the efforts that practitioners in the field are making to establish Artificial Life as a valid, significant, and exciting area of scientific research. These efforts are not unrelated to the visual and verbal transformations discussed above. To the extent that the "creatures" are biomorphized, their representation reinforces the strong claim that the "creatures" are actually alive and extends its implications. Nor do the transformations appear only in the video, although they are particularly striking there. As the discussion above demonstrates, they are also inscribed in published articles and commentary. In fact, they are essential to the strong claim that the computer codes do not merely simulate life but are themselves alive. At least some researchers at the Santa Fe Institute recognize the relation between the strong claim and the stories researchers tell about these "organisms." Asked about the strong claim, one respondent insisted, "It's in the eye of the beholder. It's not the system, it's the observer."[12]

In the second wave of cybernetics, accounting for the observer was of course a central concern. What happens when the observer is taken into account in Artificial Life research? To explore further the web of connections between the program's operations, descriptions of its operation by observers, and the contexts in which these descriptions are embedded, we will follow the thread to the next narrative level, where arguments circulate about the contributions Artificial Life can make to scientific knowledge.

## Positioning the Field: The Politics of Artificial Life

Christopher Langton, one of the most visible of AL researchers, explains the reasoning behind the strong claim. "The principle [sic] assumption made in Artificial Life is that the 'logical form' of an organism can be separated from its material basis of construction, and that 'aliveness' will be found to be a property of the former, not of the latter."[13] It would be easy to dismiss the claim on the basis that the reasoning behind it is tautological: Langton defines life in such a way as to make sure the programs qualify, and then, because they qualify, he claims they are alive. But more is at work here than tautology. Resonating through Langton's definition are assumptions that have marked Western philosophical and scientific inquiry at least since Plato. Form can logically be separated from matter; form is privileged over matter; *form defines life*, while the material basis merely instantiates it. The definition is a site of reinscription as well as tautology. This convergence suggests that the context for our inquiry should be broadened beyond the definition's logical form to the field of inquiry in which such arguments persuade precisely because they reinscribe.

This context includes attitudes, deeply by many researchers in scientific communities, about the relation between the complexity of observable phenomena and the relatively simple rules they are seen to embody. Traditionally the natural sciences, especially physics, have attempted to reduce apparent complexity to underlying simplicity. The attempt to find the "fundamental building blocks" of the universe in quarks is one example of this endeavor; the mapping of the human genome is another.[14] The sciences of complexity, with their origins in nonlinear dynamics, complicated this picture by demonstrating that for certain nonlinear dynamical systems, the evolution of the system could not be predicted, even in theory, from the initial conditions (as Ray did not know what creatures would evolve from the ancestor). Thus the sciences of complexity articulated a limit on what reductionism could accomplish. In a significant sense, however, AL researchers have not relinquished reductionism. In place of predictability, traditionally the test of whether a theory works, they emphasize emergence. Instead of starting with a complex phenomenal world and reasoning back through chains of inference to what the fundamental elements must be, they *start* with the elements and complicate them through appropriately nonlinear processes so that the complex phenomenal world appears on its own.[15]

Why is one justified in calling the simulation and the phenomena that emerge from it a "world"? Precisely because they are generated from simple underlying rules and forms. AL reinscribes, then, the mainstream assumption that simple rules and forms give rise to phenomenal complexity. The difference is that AL starts at the simple end where synthesis can move forward spontaneously, rather than at the complex end where analysis must work backward. Christopher Langton, in his explanation of what AL can contribute to theoretical biology, makes this difference explicit. "Artificial Life," he writes, "is the study of man-made systems that exhibit behaviors characteristic of natural living systems. It complements the traditional biological sciences concerned with the *analysis* of living organisms by attempting to *synthesize* life-like behaviors within computers and other artificial media. By extending the empirical foundation upon which biology is based beyond the carbon-chain life that has evolved

on Earth, Artificial Life can contribute to theoretical biology by locating *life-as-we-know-it* within the larger picture of *life-as-it-could-be* ."[16]

The presuppositions informing such statements have been studied by Stefan Helmreich, an anthropologist who spent several months at the Santa Fe Institute.[17] Helmreich interviewed several of the major players in the American AL community, including Christopher Langton and Thomas Ray, about whom we have already heard, as well as John Holland and others. He summarizes the views of his informants about the "worlds" they create. "For many of the people I interviewed, a 'world' or 'universe' is a self-consistent, complete, and closed system that is governed by low level laws that in turn support higher level phenomena which, while dependent on these elementary laws, cannot be simply derived from them."[18] Helmreich uses comments from the interviews to paint a fascinating picture about the various ways in which simple laws are believed to underlie complex phenomena. Several informants thought that the world was mathematical in essence. Others held the view, also extensively articulated by Edward Fredkin (about whom we will hear more shortly), that the world is fundamentally comprised of information.[19] From these points of view, phenomenological experience is itself a kind of illusion, covering over an underlying reality of simple forms. For them, a computer program that generates phenomenological complexity out of simple forms is no more or less illusory than the "real" world.

The form/matter dichotomy is intimately related to this vision, for reality at the fundamental level is seen as form rather than matter, specifically as informational code whose essence lies in a binary choice rather than a material substrate. Fredkin, for example, says that reality is a software program run by a cosmic computer, whose nature must forever remain unknown to us because it lies outside the programs that run on it.[20] For Fredkin, AL programs are alive in precisely the same sense as biological life--because they are complex phenomena generated by underlying binary code. The assumption that form occupies a foundational position relative to matter is especially easy to make with information technologies, since information is defined in theoretic terms (as we have seen) as a probability function and thus as a pattern or form rather than a materially instantiated entity.

Information technologies seem to realize a dream impossible in the natural world--the opportunity to look directly into the inner workings of reality at its most elemental level. The directness of the gaze does not derive from the absence of mediation. On the contrary, our ability to look into programs like Tierra is highly mediated by everything from computer graphics to the processing program that translates machine code into a high-level computer language such as C++. Rather, the gaze is privileged because the observer can peer directly into the elements that the world is before it cloaks itself with the appearance of complexity. Moreover, the observer is presumed to be cut from the same cloth as the world he inspects, inasmuch as he is also constituted through binary processes similar to those he sees inside the computer. The essence of Tierra as an artificial world is no different from the essence of the observer or the world he occupies: all are constituted through forms understood as informational patterns. When *form* is triumphant, Tierra's "creatures" are, in a disconcertingly literal sense, just as much *life forms* as any other organisms.

We are now in a position to understand the deep reasons why some practitioners think of programs like Tierra not as models or simulations but life itself. As Langton and many others point out, in the analytic approach reality is modeled by treating a complex phenomenon as if it were comprised of smaller constituent parts. These parts are broken down into still smaller parts, until one arrives at parts sufficiently simplified so that they can be treated mathematically. Most scientists would be quick to agree that the model is not the reality, because they recognize that many complexities had to be tossed out by the wayside in order to lighten the wagon sufficiently to get it over the rough places in the trail. Their hope is that the model nevertheless captures enough of the relevant aspects of a system to tell them something significant about how reality works. In the synthetic approach, by contrast, the complexities emerge spontaneously as a result of the system's operation. The system itself adds back in the baggage that had to be tossed out in the analytic approach. (Whether it is the same baggage remains, of course, to be seen.) In this sense Artificial Life poses an interesting challenge to the view of nineteenth-century vitalists, who saw in the analytic approach a reductionist methodology that could never adequately capture the complexities of life. If it is true that the analytical approach murders by dissection, by the same reasoning the synthetic approach of AL may be able to procreate by emergence.

In addition to these philosophical considerations, there are also more obviously political reasons to make a strong claim for the "aliveness" of AL. As a new kid on the block, Artificial Life must jockey for position with larger, more well-established research agendas. A common reaction from other scientists is, "Well, this is all very interesting, but what good is it?" Even AL researchers joke that AL is a solution in search of a problem. When applications are suggested, they are often open to cogent objections. As long as AL programs are considered to be simulations, any results produced from them may be artifacts of the simulation rather than properties of

natural systems. So what if a certain result can be produced within the simulation? It is artifactual and therefore non-signifying with respect to the natural world unless the same mechanisms can be shown to be at work in natural systems.[21] These difficulties disappear, however, if AL programs are themselves alive. Then the point is not that they model natural systems but rather that they are, *in themselves*, also alive and therefore as worthy of study as evolutionary processes in naturally-occurring media.

This is the tack that Christopher Langton takes when he compares AL simulations to synthetic chemicals.[22] In the early days, he observes, the study of chemistry was confined to naturally occurring elements and compounds. Although some knowledge could be gained from these, the results were limited by what lay ready at hand. Once researchers learned to synthesize chemicals, their knowledge took a quantum leap forward, for then chemicals could be tailored to specific research problems. Similarly, theoretical biology has been limited to the case that lay ready to hand, namely the evolutionary pathways taken by carbon-based life. It is notoriously difficult to generalize from a single instance, but theoretical biology had no choice; carbon-based life was it. Now a powerful new instance has been added to the repertoire, for AL simulations represent an alternative evolutionary pathway followed by silicon-based life forms.

What theoretical biology looks for, in this view, are similarities that cut across the particularities of the media. In "Beyond Digital Naturalism," Walter Fontana and his coauthors lay out a research agenda "ultimately motivated by a premise: that there exists a logical deep structure of which carbon chemistry-based life is a manifestation. The problem is to discover what it is and what the appropriate mathematical devices are to express it".[23] Such a research agenda presupposes that the essence of life, understood as logical form, is independent of the medium. More is at stake in this agenda than expanding the frontiers of theoretical biology. By positing AL as a second instance of life, researchers affect the definition of biological life as well, for now it is the juxtaposition that determines what counts as fundamental, not carbon-based forms by themselves. This change hints at how far-reaching the implications can be of the narrative of Artificial Life as an alternate evolutionary pathway for life on earth. To explore these implications, let us turn to the third level of narrative, where we will consider stories about the relation of humans to our silicon cousins, the Artificial Life forms who represent the road not taken--until now.

## Reconfiguring the Body of Information

As research on Artificial Life forms continues and expands, the construction of *human* life is affected as

well. Two different narratives of how the human will be reconfigured in the face of artificial bodies of information are told by Rodney Brooks of the Artificial Intelligence Laboratory at MIT, and Hans Moravec, of whom we have already heard. Whereas Moravec privileges consciousness as the essence of human being and wants to preserve it intact, Brooks speculates that the more essential property is the ability to move around and interact robustly with the environment. Instead of starting with the most advanced qualities of human thought, Brooks starts with locomotion and simple interactions and works from the bottom up. Despite these different orientations, both Brooks and Moravec see the future of human being inextricably bound up with Artificial Life. Indeed, in the future world they envision, it will be difficult or impossible to distinguish between natural and artificial life, human and machine intelligence.

In *Mind Children: The Future of Robot and Human Intelligence,* Moravec argues that the age of carbon-based life is drawing to a close.[24] Humans are about to be replaced as the dominant life form on the planet by intelligent machines. Drawing on the work of Cairns-Smith, Moravec suggests that such a revolution is not unprecedented. Before protein replication developed, a primitive form of life existed in certain silicon crystals that had the ability to replicate. But protein replication was so far superior that it soon left the replicating crystals in the dust. Now silicon has caught up with us again, in the form of computers and computerized robots. Although the Cairns-Smith hypothesis has been largely discredited, in Moravec's text it serves the useful purpose of increasing the plausibility of his vision by presenting the carbon-silicon struggle as a rematch of a earlier contest rather than an entirely new event.

A different approach is advocated by other members of the Artificial Life community, among them Rodney Brooks, Pattie Maes, and Mark Tilden.[25] They point to the importance of having agents who can learn from interactions with a physical environment. Simulations, they believe, are limited by the artificiality of their context. Compared to the rich variety and creative surprises of the natural world, simulations are stick worlds populated by stick figures. No one argues this case more persuasively than Brooks. When I talked with him at his MIT laboratory, he mentioned that he and Hans Moravec were roommates in college (a coincidence almost allegorical in its neatness). Moravec, for his senior project, had built a robot that used a central representation of the world to navigate. The robot would go a few feet, feed in data from its sensors to the central representation, map its new position, and move a few more feet. Using this process, it would take several hours to cross a room. If anyone came in during the meantime, it would be thrown hopelessly

off. Brooks, a loyal roommate, stayed up late one night to watch the robot as it carried out its agonizingly slow perambulation. It occurred to him that a cockroach could accomplish the same task in a fraction of the time, and yet the cockroach could not possibly have as much computing power aboard as the robot. He decided that there had to be a better way and began building robots according to a different philosophy.

In his robots, Brooks uses what he calls "subsumption architecture." The idea is to have sensors and actuators connected directly to simple finite state machine modules, with a minimum of communication between them. Each system "sees" the world in an entirely different way from the others. There is no central representation, only a control system that kicks in to adjudicate when there is a conflict between the distributed modules. Brooks points out that the robot does not need to have a coherent concept of the world; instead it can learn what it needs directly through interaction with its environment. The philosophy is summed up in his aphorism: "the world is its own best model."[26]

Subsumption architecture is designed to facilitate and capitalize on emergent behavior. The idea can be illustrated with Genghis, a six-legged robot somewhat resembling a oversize cockroach, that Brooks hopes to sell to NASA as a planetary explorer.[27] Genghis's gait is not programmed in advance. Rather, each of the six legs is programmed to stabilize itself in an environment that includes the other five. Each time Genghis starts up, it has to learn to walk anew. For the first few seconds it will stumble around; then, as the legs begin to take account of what the others are doing, a smooth gait emerges. The robot is relatively cheap to build, more robust than the large planetary explorers NASA currently uses, and under its own local control rather than dependent on a central controller who may not be on site to see what is happening. "Fast, cheap, and out of control" is another aphorism that Brooks uses to sum up the philosophy behind the robots he builds.

Brooks's program has been carried further by Mark Tilden, a Canadian roboticist who worked under Brooks and now is at the University of Waterloo. In my conversation with him, Tilden mentioned that he grew up on a farm in Canada and was struck by how chickens ran around after they had their heads cut off, performing, as he likes to put it, complicated navigational tasks in three-dimensional space without any cortex at all. He decided considerable computation had to be going on in the peripheral nervous system. He used the insight to design insect-like robots that operate on nervous nets (considerably simpler than the more complex neural nets) comprised of no more than 12 transistor circuits. These robots use analogue rather than

digital computing to carry out their tasks. Like Genghis, their gait is emergent. They are remarkably robust, able to right themselves when turned over, and can even learn a compensatory gait when one of their legs is bent or broken off. [28]

Narratives about the relation of these robots to humans emerge when Brooks and others speculate about the relevance of their work to human evolution. Brooks acknowledges that the robots he builds have the equivalent of insect intelligence. But insect intelligence is, he says, nothing to sneer at. Chronologically speaking, by the time insects appeared on earth, evolution was already 95% of the way to creating human intelligence.[29] The hard part, he believes, is evolving creatures who are mobile and who can interact robustly with their environment. Once these qualities are in place, the rest comes relatively quickly, including the sophisticated cognitive abilities that humans possess. How did humans evolve? In his view, through the same kind of mechanisms that he uses in his robots, namely distributed systems that interact robustly with the environment and that consequently "see" the world in very different ways. Consciousness is a relatively late development, analogous to the control system that kicks in to adjudicate conflicts between the different distributed systems. Consciousness is, as Brooks likes to say, a "cheap trick," that is, an emergent property that increases the functionality of the system but is not part of system's essential architecture. Consciousness does not need to be, and in fact is not, representational. Like the robot's control system, consciousness does not require an accurate picture of the world; it only needs a reliable interface. As evidence that human consciousness works this way, Brooks adduces the fact that most adults are unaware that they go through life with a large blank spot in the middle of their visual field.

This reasoning leads to yet another aphorism that circulates through the AL community: consciousness is an epiphenomenon. The implication is that consciousness, although it thinks it is the main show, is in fact a latecomer, a phenomenon dependent on and arising from deeper and more essential layers of perception and being. The view is reminiscent of the comedian Emo Phillips's comment. "I used to think that the brain was the most wonderful organ in the body," he says. "But then I thought, who's telling me this?"

It would be difficult to imagine a more contrarian position to the one Hans Moravec espouses when he equates consciousness with human subjectivity. In this respect Moravec aligns with artificial intelligence (AI), whereas Brooks and his colleagues align with Artificial Life (AL).[30] Michael Dyer, in his comparison of the two fields, points out that whereas AI envisions cognition as the operation of

logic, AL sees it as the operation of nervous systems; AI starts with human-level cognition, AL with insect- or animal-level; in AI, cognition is constructed as if it were independent of perception, whereas in AL it is integrated with sensory/motor experiences.[31]

Brooks and his colleagues argue forceful that AI has played itself out and that the successor paradigm is AL. Brooks and Ray both believe that it will eventually be possible, using AL techniques, to evolve the equivalent of human intelligence inside a computer. For Brooks, that project is already underway with "Cog," a head-and-torso robot with sophisticated visual and manipulative capability. But AL researchers go about creating high-level intelligence in dramatically different ways that AI. Consider the implications of this shift for the construction of the human. The goal of artificial intelligence was to build an intelligence inside a machine comparable to that of a human. The human was the measure, the machine the attempt at instantiation in a different medium. This assumption deeply informs the Turing test, dating from the early days of the AI era, which defined success as building a machine intelligence that cannot be distinguished from a human intelligence. By contrast, the goal of Artificial Life is to evolve intelligence within the machine through pathways found by the "creatures" themselves. Rather than serving as the measure to judge success, human intelligence is itself is reconfigured in the image of this evolutionary process. Whereas AI dreamed of creating consciousness inside a machine, AL sees human consciousness, understood as an epiphenomenon, perching on top of the machine-like functions that distributed systems carry out.[32] In the AL paradigm, the machine becomes the model for understanding the human. Thus the human is transfigured into the posthuman.

To indicate how widespread is this refashioning of the human into the posthuman, in the following section I want to sketch with broad strokes some of the research contributing to this project. The sketch will necessarily be incomplete. Yet even this imperfect picture will be useful in indicating the scope of the posthuman. So pervasive is this refashioning that it amounts to a new world view-- still in process, highly contested and often speculative, yet with enough links between different sites to be edging toward a vision of what we might call the computational universe. In the computational universe, the essential function for both intelligent machines and humans is processing information. Indeed, the essential function of the universe as a whole is processing information. In a different way than Norbert Wiener imagined, the computational universe realizes the cybernetic dream of creating a world in which humans and intelligent machines can both feel at home. That equality derives from the view that our world--and the great cosmos itself--is a vast computer, and we are programs it runs.

## The Computational Universe

Let us start our tour of the computational universe at the most basic level, the level that underlies all life forms, indeed all matter and energy. The units that comprise this level are cellular automata. From their simple on-off functioning, everything else is built up. Cellular automata were first proposed by John von Neumann in his search to describe self-reproducing automata. Influenced by Warren McCulloch and Walter Pitts's work on the on-off functioning of the neural system, von Neumann used the McCulloch-Pitts neuron as a model for computers, inventing switching devices that could perform the same kind of logical functions McCulloch had outlined for neurons. He also proposed that the neural system could be treated as a Turing machine. Biology thus provided him with clues to build computers, and computers provided clues for theoretical biology. To extend the analogy between biological organism and machine, he imagined a giant automaton that could perform the essential biological function of self-reproduction.[33] (Maturana referred to it, as we saw in Chapter 6, when he made the point that what von Neumann modeled were biologists' descriptions of living processes rather than the processes themselves.)

Stanislaw Ulam, a Polish mathematician who worked with von Neumann at Los Alamos during World War II, suggested to von Neumann that he could achieve the same result by abstracting the automaton into a grid of cells. Thus von Neumann reduced the massive and resistant materiality of the self-reproducing automaton as he had originally envisioned it to undifferentiated cells with bodies so transparent they were constituted as squares marked off on graph paper and later as pixels on computer screens.[34]

Each cellular automaton (or CA) functions as a simple finite state machine, with its state determined solely by its initial condition (on or off), rules telling it how to operate, and the state of its neighbors at each moment. For example, the rule for one group of CAs might state "On if two neighbors are on, otherwise off." Each cell checks on the state of its neighbors and updates its state in accordance with its rules, at the same time that the neighboring cells also update their states. In this way the grid of cells goes through one generation after another, in a succession of states that (on a computer) can easily stretch to hundreds of thousands of generations. Extremely complex patterns can build up, emerging spontaneously from interactions between the CAs. Programmed into a computer and displayed on the screen, CAs give the uncanny impression of being alive. Some patterns spread until they look like the

designs of intricate Oriental carpets, others float across the screen like gliders, and still others flourish only to die out within a few hundred generations. Looking at the emergence of complex dynamical patterns from these simple components, more than one researcher has had the intuition that such a system can explain the growth and decay of patterns in the natural world. Edward Fredkin took this insight further, seeing in cellular automata the foundational structure from which everything in the universe is built up.

How does this building up occur? In the computational universe, the question can be re-phrased by asking how higher-level computations can emerge spontaneously from the underlying structure of cellular automata. Christopher Langton has done pioneering work analyzing the conditions under which cellular automata can support the fundamental operations of computation, which he analyzes as requiring the transfer, storage, and modification of information.[35] His research indicates that computation is most likely to arise at the boundary between ordered structures and chaotic areas. In an ordered area, the cells are tightly tied together through rules that make them extremely interdependent; it is precisely this interdependence that leads to order. But the tightly ordered structure also means that the cells as an aggregate will be unable to perform some of the essential tasks of higher-level computation, particularly the transfer and modification of information. In a chaotic area, by contrast, the cells are relatively independent of one another; this independence is what makes them appear disordered. While this state lends itself to information transfer and modification, here the storage of information is a problem, because no pattern persists for long. Only in boundary areas between chaos and order is there the necessary tension between innovation and replication that allows patterns to build up, be modified, and travel over long distances without dying out.

These results are strikingly similar to those discovered by Stuart Kauffman in his work on the origins of life. Kauffman was Warren McCulloch's last protégé; McCulloch said in several interviews that he regarded Kauffman as his most important collaborator since Walter Pitts.[36] Kauffman argues that natural selection alone is not enough to explain the relatively short time scale on which life arose.[37] Some other ordering principle is necessary, which he locates in the ability of complex systems spontaneously to self-organize. Calculating the conditions necessary for large molecules to organize spontaneously into the building blocks of life, he found that life is most likely to arise at the edge of chaos. This means that there is a striking correspondence between the conditions under which life is likely to emerge and those under which

computation is likely to emerge--a convergence regarded by many researchers as an unmistakable sign that computation and life are linked at a deep level. In this view, humans are programs that run on the cosmic computer. When humans build intelligent computers to run AL programs, they replicate in another medium the same processes that brought themselves into being.

An important reason why such connections can be made so easily between one level and another is that in the computational universe, everything is reducible at *some* level to information.

Yet among proponents of the computational universe, not everyone favors disembodiment, just as they did not in the Macy conferences when the idea of information was being formulated. Consider, for example, the different approaches taken by Edward Fredkin and the new field of evolutionary psychology. When Fredkin asserts that we can never know the nature of the cosmic computer on which we run as programs, he puts the ultimate material embodiment out of our reach. All we will ever see, as human beings, are the informational forms of pure binary code that he calls cellular automata. By contrast, the field of evolutionary psychology seeks to locate modular computer programs in *embodied* human beings whose physical make-up is the result of hundreds of thousands of years of evolutionary processes.

The agenda for this new field is set out by Jerome H. Barkow, Leda Cosmides and John Tooby in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Like Minksy, they argue that the model (or metaphor) of computation provides the basis for a wholesale revision of what counts as human nature.[38] They aim to overcome objections by cultural anthropologists and others to the idea of "human nature" by offering a more flexible version of how that nature is constituted. They argue that behavior can be modeled as modular computer programs running in the brain. The underlying structure of these programs is the result of thousands of years of evolutionary tinkering. Adaptations which conferred superior reproductive fitness survived; those that did not died out. The programs are structured to enable certain functionalities to exist in humans, and these functionalities are universally present in all humans. These functionalities, however, represent potentials rather than actualities. Just as the actual behavior of a computer program is determined by a constant underlying structure and varying inputs, so actual human behavior results from an interplay between the potential represented by the functionalities and inputs provided by the environments. All normal human infants, for example, have the potential to learn language. If they are not exposed to language by a certain critical age, however, this potential

disappears and they can never become linguistically competent. Although human behavior varies across a wide spectrum of actualization, it nevertheless has an underlying universal structure determined by evolutionary adaptations. Thus a science of evolutionary psychology is possible, for the existence of a universal underlying structure guarantees the regularities that any science needs to formulate knowledge that will be coherent and consistent.

This cybernetic/computer vision of human behavior leads to a very different account of "human nature." Although the evolutionary programs the brain/computer runs do not lead to universal behavior, they are nevertheless rich with content. The potentials lie not just in the structure of the general machine but much more specifically in environmentally adaptive programs that proactively shape human responses. Thus children are not merely capable of learning language; they actively *want* to learn language and will invent it among themselves if no one teaches them.[39] Like Wiener's cybernetic machine, the cybernetic brain is responsive to the flow of events around it and adaptive over an astonishingly diverse set of circumstances. It is a measure of how much our vision of machines has changed since the Industrial Revolution that only the intelligent machine is seen to be light enough on its feet to do justice to human variability.

It will now perhaps be clear why the most prized functionality is the ability to process information, for in the computational universe, information is king. Luc Steels, an Artificial Life researcher, reinscribes this value when he distinguishes between first-order and second-order emergence (surely it is no accident that the terminology here echoes the distinction between first- and second-order cybernetics, the grandparent and parent of Artificial Life). First-order emergence denotes any property that is generated by interactions between components, that is, properties that emerge as a result of those interactions, in contrast to properties inherent in the components themselves. Among all such emergent properties, second-order emergence grants special privilege to those that bestow additional functionality on the system, particularly the ability to process information.[40] To create successful Artificial Life programs, it is not enough to create just any emergence. Rather, the programmer searches for a design that will lead to second-order emergence. Once second-order emergence is achieved, the organism has in effect *evolved the capacity to evolve*. Then evolution can really take off. Humans evolved through a combination of chance and self-organizing processes until they reached the point where they could take conscious advantage of the principles of self-organization to create evolutionary mechanisms. They used this ability to build machines capable of self-evolution. Unlike

humans, however, the machine programs are not hampered by the time restrictions imposed by biological evolution and physical maturation. They can run through hundreds of generations in a day, millions in a year. Until very recently, humans have been without peer in their ability to store, transmit and manipulate information. Now they share that ability with intelligent machines. To foresee the future of this evolutionary path, we have only to ask which of these organisms, competing in many ways for the same evolutionary niche, has the information-processing capability to evolve more quickly.

This conclusion makes clear, I think, why the computational universe should not be accepted uncritically. If the name of the game is processing information, it is only a matter of time until intelligent machines replace us as our evolutionary heirs. Whether we decide to fight them or join them by becoming computers ourselves, the days of the human race are numbered. The problem here does not lie in the choice between these options; rather, it lies in the framework constructed so as to make these options the only two available. The computational universe becomes dangerous when it goes from being a useful heuristic to an ideology that privileges information over everything else. As we have seen, information is a socially constructed concept that could have been, and was, envisioned differently than its currently accepted definition. Just because information has lost its body does not mean that humans or the world have lost theirs.

Fortunately, not all theorists agree that it makes sense to think about information as an entity apart from the medium that embodies it. Let us re-visit some of the sites in the computational universe, this time to locate those places where the resistance of materiality does useful work within the theories. From this perspective, fracture lines appear that demystify the program(s) and make it possible to envision other futures than the one sketched above, futures in which human beings feel at home in the universe because they are embodied creatures living in an embodied world.

## Murmurs from the Body

One of the striking differences between researchers who work with flesh and those who work with computers is how nuanced the sense of the body's complexity is for those who are directly engaged with it. The difference can be seen in the contrast between Marvin Minsky's "Society of Mind" approach and that of the evolutionary psychologists. Although Minsky frequently uses evolutionary arguments to clarify a program's structure, his main interest clearly lies in building computer models that can accomplish human behaviors.[41] He characteristically thinks in terms of computer architecture, about which he knows a great deal,

rather than human physiology. In his lectures (less so in his writing), he rivals Moravec in his consistent downplaying of the importance of embodiment. At the public lecture he delivered on the eve of the Fifth Conference on Artificial Life in Nara, Japan, he argued that only with the advent of computer languages has a symbolic mode of description arisen adequate to account for human beings, whom he defines as complicated machines.[42] "A person is not a head and arms and legs," he remarked. "That's trivial. A person is a very large multiprocessor with a million times a million small parts, and these are arranged as a thousand computers." It is not surprising, then, that he shares with Moravec the dream of banishing death by downloading consciousness into a computer. "The most important thing about each person is the data, and the programs in the data that are in the brain. And some day you will be able to take all that data, and put it on a little disk, and store it for a thousand years, and then turn it on again and you will be alive in the fourth millennium or the fifth millennium."

Yet anyone who actually works with embodied forms, from the relatively simple architecture of robots to the vastly more complicated workings of the human neural system, knows that it is by no means trivial to deal with the resistant materialities of embodiment. To Minsky, these problems of embodiment are nuisances that do not even have the virtue of being conceptually interesting. In his plenary lecture at Artificial Life V, he asserted that a student who constructed a simulation of robot motion learned more in six months than the roboticists did in six years from building actual robots.[43] Certainly simulations are useful for a wide range of problems, for they abstract a few features out of a complex interactive whole and then manipulate those features to get a better understanding of what is going on. Compared to the real world, they are more efficient precisely because they are more simplified. The problem comes when this mode of operation is taken to be fully representative of a much more complex reality, and everything that is not in the simulation is declared to be trivial, unimportant, or uninteresting.

Like Varela in his criticisms and modifications of Minsky's model (discussed in Chapter 6), Barkow, Tooby and Cosmides are careful not to make this mistake. They acknowledge that the mind-body duality is a social construction which obscures the holistic nature of human experience. Another researcher who speaks powerfully to the importance of embodiment is Antonio Damasio in *Descartes' Error: Emotion, Reason, and the Human Brain*.[44] Discussing the complex mechanisms by which mind and body communicate, he emphasizes that the body is more than a life support system for the brain. The body "contributes a content that is part and parcel of the workings of the normal mind" (p. 226). Drawing

upon his detailed knowledge of neurophysiology and his years of experience working with patients who have suffered neural damage, he argues that feelings constitute a window through which the mind looks into the body. Feelings are how the body communicates to the mind information about its structure and continuously varying states. If feelings and emotions are the body murmuring to the mind, then feelings are "just as cognitive as other precepts," part of thought and indeed part of what makes us rational creatures (p. xv). He finds it significant that cognitive science, with its computational approach to mind, has largely ignored the fact that feelings even exist (with some notable exceptions such as *The Embodied Mind*, discussed in Chapter 6). One can guess what his response to the scenario of downloading human consciousness would be from the following passage. "In brief, neural circuits represent the organism continuously, as it is perturbed by stimuli from the physical and sociocultural environments, and as it acts on those environments. If the basic topic of those representations were not an organism anchored in the body, we might have some form of mind, but I doubt that it would be the mind we do have" (226). Human mind without human body is not human mind. More to the point, it doesn't exist.

What are we to make, then, of the posthuman? As the liberal humanist subject is dismantled, many parties are contesting to determine what will count as (post)human in its wake. For most of the researchers discussed in this chapter, becoming a posthuman means much more than having prosthetic devices grafted onto one's body. It means envisioning humans as information-processing machines with fundamental similarities to other kinds of information-processing machines, especially intelligent computers. Because of how information has been defined, many people holding this view tend to put materiality on one side of a divide and information on the other side, making it possible to think of information as a kind of immaterial fluid that circulates effortlessly around the globe while still retaining the solidity of a reified concept. Yet this is not the only view, and in my judgment, it is not the most compelling one. Other voices insist that the body cannot be left behind, that the specificities of embodiment matter, that mind and body are finally the "unity" Maturana insisted upon rather than two separate entities. Increasingly the question is not whether we will become posthuman, for posthumanity is already here. Rather, it is what kind of posthumans we will be. What the narratives of Artificial Life reveal is that if we acknowledge that the observer *must* be part of the picture, bodies can never be made of information alone, no matter which side of the computer screen they are on.

## Endnotes

[1] Francisco Varela, Evan Thompson and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge: MIT Press, 1991).

[2]Francisco Varela and Paul Bourgine, editors, *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life* (Cambridge: MIT Press, 1992).

[3] Humberto R. Maturana and Francisco Varela, *Autopoiesis and Cognition: The Realization of the Living* (Dordrecht: D. Reidel, 1980).

[4]Thomas S. Ray, "A Proposal to Create Two Biodiversity Reserves: One Digital and One Organic," presentation at Artificial Life IV, Cambridge MA, July 1994.

[5]Thomas S. Ray, "An Evolutionary Approach to Synthetic Biology: Zen and the Art of Creating Life," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): 179-209, especially 180.

[6]Luc Steels offers useful definitions of emergence in "The Artificial Life Roots of Artificial Intelligence," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): 75-110. He distinguishes between first-order emergence, defined as a property not explicitly programmed in, and second-order emergence, an emergent behavior that adds additional functionality to the system. In general AL researchers try to create second-order emergence, for then the system can use its own emergent properties to create an upward spiral of continuing evolution and emergent behaviors. James P. Crutchfield makes a similar point in "Is Anything Ever New? Considering Emergence," in *Integrative Themes*, edited by G. Cowan, D. Pines, and D. Melzner, Santa Fe Institute Studies in the Sciences of Complexity XIX (Redwood City CA: Addison-Wesley, 1994): 1-15. For a criticism of emergence, see Peter Cariani, "Adaptivity and Emergence in Organisms and Devices," *World Futures* 32 (1991): 111-32.

[7]The Tierra program is described in Thomas S. Ray, "An Approach to the Synthesis of Life," *Artificial Life II*, edited by Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, Proceedings Volume X, Santa Fe Institute Studies in the Sciences of Complexity (Redwood City CA: Addison-Wesley, 1992), pp. 371-408. "An Evolutionary Approach to Synthetic Biology" explains and expands on the philosophy underlying Tierra, working paper, ATR Human Information Processing Research Laboratories, Kyoto Japan. Further information about Tierra can be found in "Population Dynamics of Digital Organisms," *Artificial Life II Video Proceedings*, edited by Christopher G. Langton (Redwood City CA: Addison-Wesley, 1991). A popular account can be found in "Electronic Ecosystem," John Travis, *Science News* 140, no. 6 (August 10, 1991): 88-90.

[8]"Simple Rules . . . Complex Behavior," produced and directed by Linda Feferman for the Santa Fe Institute, 1992.

[9]Richard Dawkins, *The Selfish Gene* (Oxford: Oxford University Press, first edition 1976).

[10]William Gibson, *Neuromancer* (New York: Ace, 1984).

[11]In "An Evolutionary Approach," Ray writes that "The 'body' of a digital organism is the information pattern in memory that constitutes its machine language program" (184).

[12]Quoted in Stefan Helmreich, "Anthropology Inside and Outside the Looking-Glass Worlds of Artificial Life," unpublished manuscript (1994), p. 11. An earlier version of this work was published as a working paper at the Santa Fe Institute under the title "Travels through 'Tierra,' Excursions in 'Echo': Anthropological Refractions on the Looking-Glass Worlds of Artificial Life," 94-04-024. Helmreich included in this version some remarks that the administrators of SFI evidently found offensive, including a comparison between the scientists' belief in the "aliveness" of Artificial Life to the seemingly bizarre (to Westerners) beliefs held by such marginal cultural groups as the Trobriand Islanders. Objecting that Helmreich's work was not scientific and misrepresented the science done at SFI, the administrators had the working paper removed from the shelves and deleted from the list of available publications.

[13]Christopher Langton, "Artificial Life," in *Artificial Life*, edited by Christopher Langton (Redwood City CA: Addison-Wesley, 1989), pp. 1-47, especially p. 1.

[14]Richard Doyle has written on the simplification of body to information in the human genome project in *On Beyond Living:*

*Rhetorics of Vitality and Post-Vitality in Molecular Biology* (Stanford: Stanford University Press, 1997).

[15]Actually, both inference and deduction are at work in most AL research, as they usually are in scientific projects. AL researchers study the complex--> simple route for clues on how to construct programs that will be able to move from simple--> complex.

[16]Christopher Langton, "Artificial Life," p. 1.

[17]Extensive interviews with AL researchers have also been conducted by Steven Levy, as recounted in his useful popularization, *Artificial Life: The Quest for a New Creation* (New York: Pantheon Books, 1992). A more technical account covering much the same material as Levy can be found in Claus Emmeche, *The Garden in the Machine: The Emerging Science of Artificial Life* (Princeton: Princeton University Press, 1994). Although Emmeche says in the opening pages that his book is intended for the general reader, he soon leaves the simplistic style that characterizes the first sections and moves into more interesting and demanding material. Especially noteworthy is his discussion of the deep problems raised about the nature of computation.

[18]Helmreich, p. 5.

[19]Edward Fredkin is something of a cult figure for researchers interested in computational philosophies. After achieving financial independence through the company he founded, he bought and occasionally lives on his own island in the Caribbean. Although he has himself published very little, several articles and part of a book have been written about him. He is a faculty member at MIT and has a research group there working out a universal theory of cellular automata intending to show how cellular automata can account for all the laws of physics. See for an account of his work, see Robert Wright, *Three Scientists and Their Gods*: *Looking for Meaning in an Age of Information* (New York: Times Books, 1988). One of Fredkin's rare publications is "Digital Mechanics: An Information Process Based on Reversible Universal Cellular Automata," *Physica D* 45 (1990): 254-70. See also Julius Brown, "Is the Universe a Computer?", *New Scientist* 14 (July 1990): 37-39. Levy and Lemmeche both mention Fredkin.

[20]G. Kampis and V. Csanyi in "Life, Self-Reproduction and Information: Beyond the Machine Metaphor," *Journal of Theoretical Biology* 148 (1991): 17-32 have an important analysis of the idea of self-reproduction in a machine context. They point out that one's account of what happens in self-reproduction changes depending on how the framing context is constructed. For *all* machine (re)production, there is always a context in which outside agency is needed to complete reproduction, in contrast to the reproduction of [asexual] living organisms. By placing the last computer out of sight, as it were, Fredkin has erased this context from view, although he still has to posit it to explain how things come into existence.

[21]For a research program that takes this objection into account, see David Jefferson *et al.*, "Evolution as a Theme in Artificial Life: The Genesys/Tracker System," Computer Science Department Technical Report CSD-900047 (December 1990), University of California-Los Angeles. In the Tracker simulation, designed to generate social behavior and food gathering strategies characteristic of ants, Jefferson and his colleagues used two very different kinds of algorithms to demonstrate that the behavior generated by the simulations were not artifacts. They reasoned that because the underlying structures of the simulations were different, similarities of behavior could not be attributed to the algorithms, only to the dynamics conceptualized through those algorithms.

[22]Christopher Langton, "Editor's Introduction," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): v-viii, especially v-vi.

[23]Walter Fontana, Gunter Wagner, and Leo W. Buss, "Beyond Digital Naturalism," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): 211-227, especially 224.

[24]Hans Moravec, *Mind Children: The Future of Robot and Human Intelligence* (Cambridge: Harvard University Press, 1988).

[25]See Pattie Maes, "Modeling Adaptive Autonomous Agents," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): 135-162; Rodney Brooks, "New Approaches to Robotics," *Science* 253 (13 September 1991): 1227-1232; Mark Tilden, "Living Machines--Unsupervised Work in Unstructured Environments," Los Alamos National Laboratory, CB/MT-v1941114.

[26]Rodney A. Brooks, "Intelligence Without Representation," *Artificial Intelligence* 47 (1991): 139-159. See also *The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated*

*Agents*, edited by Luc Steels and Rodney Brooks (Hillsdale, NJ: L. Erlbaum Associates, 1995).

[27]Genghis is described, among other places, in Rodney A. Brooks and Anita M. Flynn, "Fast, Cheap, and Out of Control: a Robot Invasion of the Solar System," *Journal of the British Interplanetary Society*, 42 (1989): 478-485.

[28]Mark Tilden lectured and demonstrated his mobile robots at the Center for the Study and Evolution of Life, University of California, January 1995, where I had an opportunity to talk with him.

[29]Brooks, "Intelligence without Representation."

[30]In my conversation with Moravec at the University of Illinois "Cyberfest" in March 1997, he defended the top-down approach by comparing the success of a robot-piloted car he designed with Rodney Brooks' robots. While Moravec's robot car successfully drove several hundred miles, Brooks' robots have scarcely been out of the laboratory. The point is well taken, and future research may well use a combination of both approaches. Moravec declared himself a pragmatist, willing to use whatever works.

[31]Michael G. Dyer, "Toward Synthesizing Artificial Neural Networks that Exhibit Cooperative Intelligent Behavior: Some Open Issues in Artificial Life," *Artificial Life* I, no. 1/2 (Fall 1993/Winter 1994): 111-135, especially p. 112.

[32] Edwin Hutchins demystifies this proposition in *Cognition in the Wild* (Cambridge: Mit Press, 1996), when he elegantly demonstrates that humans <u>normally</u> act in environments where cognition is distributed among a variety of human and non-human actors, from graph paper and pencil to the sophisticated naval guidance systems he discusses. His book, by grounding its arguments in existing naval navigation techniques of the past and present, shows that distributed cognition has been around about as long as humans have.

[33]Steven Levy gives an account of von Neumann's self-reproducing machine in *Artificial Life*. His account is based on the rather sketchy information given by Arthur W. Burks (who edited and compiled von Neumann's incomplete manuscript after his death) of what Burks calls the kinematic model of self-reproduction. Burks' version can be found in John von Neumann, *Theory of Self-Reproducing Automata* (Urbana: University of Illinois Press, 1966), pp. 74-90.

[34]Cellular automata are described in detail in von Neumann, pp. 91-156. See also Stephan Wolfram, one of the foremost researchers on cellular automata, in "Universality and Complexity in Cellular Automata," *Physica D* 10 (1984): 1-35, and "Computer Software in Science and Mathematics," *Scientific American* 251 (August 1984): 188-203. In these articles, Wolfram concentrates on one-dimensional CAs, where each generation appears as a line and patterns appear as the lines proliferate down the screen (or graph paper).

[35]Chris G. Langton, "Computation at the Edge of Chaos: Phase Transition and Emergent Computation," *Physica D* 42 (1990); 12-37.

[36]In Warren McCulloch's papers there is a letter of reference that McCulloch wrote for Kauffmann, Warren McCulloch Papers, American Philosophical Library, B/M139, Box 2. In several lectures and interviews that McCulloch gave in the few years before his death, he mentioned Kauffman as an important collaborator.

[37]Stuart A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (New York: Oxford University Press, 1993). See also his popularized version, *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity* (New York: Oxford University Press, 1995).

[38]Jerome H. Barkow, Leda Cosmides and John Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (New York: Oxford University Press, 1992); see especially Tooby and Cosmides, "The Psychological Foundations of Culture," pp. 19-136. Tooby and Cosmides have also been instrumental in forming the "Human Behavior and Evolution Society," which holds annual conferences centered on the ideas of evolutionary psychology. In some ways the HBES is a successor to sociobiology, although with a more flexible framework of interpretation.

[39]Steven Pinker makes this point in *The Language Instinct* (New York: W. Morrow, 1994). This model provides an interesting corrective to Maturana's largely passive model of "languaging" between "observers."

[40]Luc Steels; see note 6.

[41] Marvin Minsky, *The Society of Mind* (New York: Simon and Schuster, 1985), especially pp. 17-24.

[42]Marvin Minsky, Public Lecture, "Why Computer Science is the Most Important Thing That Has Happened to the Humanities in 5,000 Years," Nara, Japan, May 15, 1996. I am grateful to Nicholas Gessler for providing me with his transcript of the lecture.

[43]Marvin Minsky, Plenary Lecture, "How Computer Science Will Change Our Lives," Fifth Conference on Artificial Life, Nara, Japan, May 17, 1996.

44Antonio R. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain* (New York: G. P. Putnam, 1994).