from John Johnston, Machinic Life: The Lure of the Post-Natural (draft)

CHAPTER 3
Vibrant Cells: Cellular Automata, Artificial Life, Autopoiesis

Yet nature is made better by no mean
But nature makes that mean: so, over that art
Which you say adds to nature, is an art
That nature makes.
William Shakespeare, A Winter's Tale

Durham tried again, but the Autoverse was holding fast to its laws. *Red* atoms could *not* spontaneously appear from nowhere – it would have violated the cellular automaton rules. And if those rules had once been nothing but a few lines of a computer program—a program which could always be halted and rewritten, interrupted and countermanded, subjugated by higher laws—that was no longer true. Zemansky was right: there was no rigid hierarchy of reality and simulation anymore. The chain of cause and effect was a loop now—or a knot of unknown topology.
Greg Egan, Permutation City

The very idea that machines may begin to reproduce themselves disturbs our conceptual boundaries. What conditions must be filled in order to initiate such a process? In a paper entitled "The General and Logical Theory of Automata," delivered at the Hixon Symposium in 1948, John von Neumann first broached this question from a logico-mathematical point of view. Subsequently, in a series of five lectures delivered at the University of Illinois in 1949, he conceived of a general plan for a self-constructing automaton. Several years later, in "The Theory of Automata: Construction, Reproduction, Homogeneity," a manuscript begun in 1952 but never published during his lifetime, von Neumann worked out an elaborate "schema" that included detailed design criteria and specifications. Even so, it would not be until the late 1970s that Christopher Langton would bring this "schema" to fruition with the creation of what has since become known as Artificial Life. Dovetailing with and relayed by theories of self-organization and emergence in nonlinear dynamical systems being developed concurrently at the Santa Fe Institute, Artificial Life spawned and exfoliated into wondrous forms, fomenting new questions about life and its relation to information. With the birth of these forms a new discipline of research was also inaugurated, reducible to neither theoretical biology nor computer science but drawing fundamentally upon both.

As an assemblage producing both new forms of behavior and a new kind of performative discursive practice, Artificial Life challenges cultural boundaries and raises novel philosophical questions. Daniel Dennett, for example, has suggested that Artificial Life is "a new way of doing philosophy," and not simply a new object for philosophy.[1] Like its putative parent discipline, Artificial Intelligence – although the filiation is misleading in fundamental respects-- Artificial Life can be viewed as the creation and testing of thought experiments, with the difference that the experimenter is kept honest by the fact that the experiments are, as Dennett puts it, "prosthetically controlled," presumably by the "simulational virtuosity of computers." Although Dennett remains tacitly coy about these important matters, he understands that Artificial Life is no ordinary science, since it creates both a new kind of object and a new methodology. One might say that its object is no longer nature but the simulation of "natural" processes that exhibit life-like behavior, but this formulation hardly conveys the extent to which hard and fast distinctions between the natural and the artificial, *phusis* and *techné*, are here undergoing radical revision and transformation.[2]

For Lacan and Freud before him, the distinction between the realm of the biological and that of the human psychic economy had been a theoretical *point d'appui* and essential support. At the same time, Lacan's introduction of the symbolic order into human life made it no longer possible to maintain any simple dualism of organism and machine. With Lacan's complex notion of "in-mixing," machines can no longer be said to reside and function outside the realm of human life, as technological extensions or modifications of it; instead, they must be understood to traverse and order it from

---

[1] Daniel Dennett, "Artificial Life as Philosophy," Artificial Life: An Overview, ed. Christopher Langton (Cambridge, Mass: The MIT Press, 1995), p. 291.

[2] One might also say that A-Life is one of the new "sciences of the artificial,'' to borrow Herbert Simon's title phrase. In The Sciences of the Artificial (first published in 1969), however, Simon is primarily concerned with questions of rationality and design; indeed, one of his chapters is called "Embedding Artifice in Nature," and relegates biology and evolutionary process to the sidelines. In contrast, John Holland's Adaptation in Natural and Artificial Systems (first published in 1975) provides not only groundbreaking conceptual tools but also a much greater foretaste of what is to come.

within. In Deleuze and Guattari's sense, these "abstract machines" operate equally in the realm of physical force and mathematical concept, in flows of both matter and information. Yet, while D&G speak of an indeterminism or free play in DNA and the genetic code (a "surplus value of code," they call it), and hence of a certain becoming-animal and a-parallel evolution, the biological realm remains largely outside their concerns, precisely because of its very capacity to reproduce and self-propagate "naturally," along lines of filiation. Against this natural mode of propagation, they extoll alliance, monstrous couplings, symbiosis, "side-communication" and "contagion," and above all those doubly deterritorializing relays they call "becomings." With the appearance of self-reproducing machines and new forms of machinic "life," however, a whole new branch of the machinic phylum begins to evolve along lines of demarcation and differentiation unlike anything given in nature. In this perspective their marginalizing of biology demands redress.[3]

Of course, the opposition between organism and machine has never been a simple empirical given; even so, this "molar" dualism has functioned as a mainstay of western metaphysics at least since Aristotle. What has enabled the boundary line between machines and biological organisms or "life" to be drawn and maintained is the basic fact that machines have never been able to reproduce themselves. In Kant's classical formulation a machine possesses only "<u>motive</u> force" [*bewegende Kraft*], in contrast with a product of nature, which is an "organized being that has within it <u>formative</u> force, and a formative force that this being imparts to the kinds of matter that lack it (thereby organizing them). This force is therefore a formative force that propagates [or reproduces] itself [*fortpflanzende bildende Kraft*] -- a force that a mere ability [of one thing] to move [another] (i.e., a mechanism) cannot explain."[4] What all

[3] See Mark Hansen's important essay, "Becoming as Creative Involution?: Contextualizing Deleuze and Guattari's Biophilosophy," in <u>Postmodern Culture</u> 11.1 (Sept. 2000). Although Hansen's purpose is to critique certain excesses in D&G's appropriation of biological theory, he also shows how closely resonant they are with aspects of contemporary biology influenced by complexity theory.

[4] Immanual Kant, <u>The Critique of Judgment</u>, trans. Werner S. Pluhar (Indianapolis: Hackett, 1987), p. 253.

machines lack (the specific example is a watch) is what Kant calls "natural purpose," which is exhibited by all "organized beings." Natural purpose, in turn, follows from two conditions, which must both be met: the parts of organized being are produced and exist for each other, and they are all part of a "self-organizing" unity of the whole. Again, machines lack this finality or purposiveness, that is, the self-organizing capacity to be self-directed; they receive not only their formal cause or purpose but also their efficient cause from outside themselves. As Kant states summarily, machines exist only "for the sake of the other."

With the advent of machinic self-organization and self-reproduction, these distinctions no longer hold. In fact, as Bernard Stiegler has pointed out, they actually begin to break down with the formation of a "dynamic of the technical system" following the rise of industrialism, and thus implicitly even before self-reproduction becomes an issue.[5] For Stiegler, who examines the concept of the "technical system" in the writings of Bertrand Gille, Gilbert Simondon, and André Leroi-Gourhan, what is at stake is the extent to which the biological concept of evolution can be applied to the technical system. In Le mode d'existence des objets techniques (1958), for example, Gilbert Simondon argues that with the Industrial Revolution a new kind of "technical object" distinguished by a quasi-biological dynamic is born. Strongly influenced by cybernetics, Simondon understands this becoming-organic of the technical object as a tendency among the systems and sub-systems that comprise it toward a unity and a constant adaptation to itself and to the changing conditions it brings about.

However, when the machines themselves begin to reproduce they attain a dramatic kind of "life" never before imagined, except under the literary or mythic aegis of the demonic and the infernal. Yet it is not from these categories and the thinking they

---

[5] See La technique et le temps, tome 1 (Paris: Galilée, 1994), esp. pp.43-94. Unfortunately Steigler doesn't consider contemporary developments like Artificial Life.

imply that machinic life will liberate itself. What we see taking place, rather, is a complex process of involution and re-articulation: not only have certain kinds of machines been constructed that can reproduce themselves, but life, biological life, has been reconceived as comprised of "living systems," that is, machines that are organized in a specific way. In 1972 the biologists Humberto Maturana and Francisco Varela proposed precisely this definition in their essay, "Autopoiesis: The Organization of the Living." Proceeding under the inspiration of cybernetic theory's "second wave," Maturana and Varela understand a living system to be a machine that not only reproduces itself but constantly regenerates itself, in all of its parts and processes, primarily through internal regulating mechanisms that maintain the organism (or machine) in a constant state of organization. They call this state "autopoiesis." Armed with this concept they have argued against the information-processing bias of molecular biology and its centralizing of DNA and cellular reproduction.

When Artificial Life and the theory of autopoiesis are considered together, as they will be in this chapter, they provide striking evidence of what might be called a double inversion, in which each side of the opposition between machines and biological organisms gives way to the other: non-organic machines become self-reproducing organisms, and organisms become autopoietic machines. Indeed, this double inversion defines an essential tendency of contemporary technological development. One ardent advocate, Kevin Kelly, marshals an impressive array of examples in order to argue that we are entering a new era defined by a "biology of machines," in which "the realm of the *born*—all that is nature—and the realm of the *made*—all that is humanly constructed—are becoming one."[6] But whether this becoming is "one" or a complex multiplicity (the position I will take), it is clear that the opposition between machine and organism no longer marks the site of a simple conceptual breach or collapse, but has become a two-

[6] Kevin Kelly, Out of Control: The Rise of Neo-biological Civilization (Reading, Mass.: Addison-Wesley, 1994), p.1. Although Kelly describes his book as "an update on the current state of cybernetic research" (453), he offers no historical perspective on the new blurring of boundaries he details.

way street or nexus out of which new conceptual possibilities and contemporary technologies are rapidly emerging.

Life on the Grid

When John von Neumann began to think about how machines might reproduce themselves, he imagined a Rube Goldberg-like apparatus floating in a lake. This apparatus, or automaton, consisted of various "organs" that would enable it to build a copy of itself out of parts freely available in the lake. The most important of these organs consisted of a "universal constructor," that is, a Turing machine with a tape control that could both store and execute instructions, as well as a constructing arm and various "elements" that could sense, cut out, and connect together other elements. In the design of the automaton certain elements called "girders" provided both structural rigidity and encoded information (i.e., their presence or absence at certain joints denoted 1s or 0s). The "lake" was simply a reservoir of elements that differed from the universal constructor only in that it was unbounded and lacked internal organization.[7]

This rather crude and unwieldy model could nonetheless serve its intended purpose, for what von Neumann sought was not so much a material device that could be constructed as a visualizable instantiation of a set of abstract logical functions. Intrinsic to the automaton's design was von Neumann's realization that in order for any entity to reproduce itself —whether that entity be natural or artificial—two separate but imbricated functions would be required: first, the entity would have to provide a blueprint with instructions that when activated would be able to direct the production of a copy of itself; and secondly, the blueprint and instructions would have to be passively copied into the offspring; otherwise, it would not be able to reproduce itself in turn. Summarily, then, the reproductive mechanism would have to contain information that could be used both

[7] I draw here on Arthur W. Burks' account in Essays on Cellular Automata, ed. Arthur C. Burks (Urbana: University of Illinois Press, 1970), 3-83, as well as von Neumann's Theory of Self-Reproducing Automata (Urbana: University of Illinois press, 1966), which Burks edited and completed.

as interpreted instructions to be executed and as passive or uninterpreted data to be copied. Watson and Crick's discovery of the structure of DNA in 1953 confirmed that nature itself utilized the same dual functionality for the reproduction of life.

Von Neumann's student Arthur Burks called this first attempt to instantiate the principles of self-reproduction in an abstract logical structure the "kinematic model." Its obvious limitation was its material complexity. So, in order to avoid unwieldy material problems like the sensing, cutting and connecting of the girders floating randomly in the lake, von Neumann decided to try a different "medium," a cellular automaton, following the suggestion of his mathematician friend Stanley M. Ulam. In its most common form a cellular automaton is a lattice or checker-board array divided into square cells, each cell of which is a finite-state automaton that "communicates" at each moment in a series of discrete time-steps with its surrounding, contiguous neighbors. At each time-step every cell takes the states of all its neighbor cells as its input; its output will then be its own state in the next time-step, as defined by a state-transition table. An example of the latter might be: if three or more of the adjacent cells are "alive" or "on," then the central cell will also be "alive" for the next time-step; if not, the cell will "die" or go quiescent.[8] In von Neumann's cellular automata there were twenty-nine possible states for each cell: one quiescent state and twenty-eight ways to be "on." Von Neumann thought that such a large number of states would be necessary in order to work out the state-transition tables that would yield a configuration of cells with a specific property, namely, that it would reproduce itself.

As with the kinematic model, the automaton was comprised of two parts: a constructor control unit and a constructor arm. Given the right programming (i.e., the

[8] With such simple rules very complex behavior can be generated, as will become evident in John Conway's "Game of Life," invented in the late 1960s following von Neumann's work. The game was popularized by Martin Gardner's "Mathematical Games: The Fantastical Combinations of John Conway's New Solitaire Game 'Life,'" Scientific American, Oct. 1970, 112-117. For a fascinating account of the game

right set of state-transition tables), the constructor control would "grow" a constructor arm out into an empty or "off" region of the cellular space. As the "head" of the constructor arm scanned back and forth, it would produce new "on" cells. Eventually a new copy of the original configuration would be built, scanned line by line and built up from the surrounding cells. Finally, when complete, it would separate itself from the original configuration and begin to reproduce itself in turn. Thus, in contrast to the kinematic model, self-reproduction in the cellular automaton takes place simply by extending into and organizing the adjacent space. As a consequence, von Neumann was able to go much further in his effort to abstract the logical form from the phenomenon of natural reproduction. For this reason, and despite the fact that von Neumann never worked out the pattern of state-transition tables (an essential but no doubt tedious task), the entire conception possesses an eerie and compelling brilliance.

After Arthur Burks had assembled and completed von Neumann's manuscripts (he also showed that in principle the model could be simulated on a computer), Edgar Codd and John Conway devised cellular automata with far fewer states per cell that could self-reproduce.[9] Then, in the late 1970s and early 80s, just as the study of cellular automata was about to lapse into complete obscurity, more physics-oriented researchers like J. Doyne Farmer, Ed Fredkin, Tommaso Toffoli, and Stephen Wolfram demonstrated how CAs could be used to model and explore dynamical systems in ways that would increase our understanding of natural phenomena.[10] Thus the work on cellular automata following von Neumann's underwent two distinct developments: the mechanism of self-reproduction was simplified, and its range of application extended. Yet there was another possibility waiting in the wings, which would eventually give birth to Artificial Life. This was the realization that cellular automata, in their capacity to self-

---

and its scientific ramifications, see William Poundstone's The Recursive Universe (Chicago: Contemporary Books, Inc., 1985).

[9] See Christopher Langton’s account in "Self-Reproduction in Cellular Automata," Physica D 10, 1-2 (1984), 135-144.

reproduce and exhibit complex behavior, suggested a form of "life" that did not simply simulate natural forms but that possessed its own ontological specificity or *quidditas*.

Christopher Langton, who had pursued the trajectory of CA research from von Neumann to Burk, Cobb and Conway more or less on his own, made this idea a concrete reality. He would later argue that biological life as we know it may be only one possible form --it is carbon-based-- of a more general process whose logic and essential properties remain to be discovered. This would mean that some of the properties of life as biologists currently study it may be only contingent properties, due solely to local conditions on earth. Working with cellular automata inevitably led Langton to the idea that life is "a property of *form*, not *matter*, a result of the organization of matter rather than something that inheres in the matter itself."[11] Neither nucleotides nor amino acids are alive, yet when combined in a certain way "the dynamic behavior that emerges out of their interactions is what we call life." Life is thus a kind of behavior, not a kind of stuff, and it is constituted from simpler behaviors. Moreover, because these behaviors arise from the nonlinear interactions among many physical parts, they include "*virtual parts*" which cease to exist when the physical parts are isolated. Artificial Life, therefore, is concerned with the "*virtual parts* of living systems." These virtual parts, in Langton's metaphor, "are the fundamental atoms and molecules of behavior" (41).[12] But while the parts may be only computer simulations, the dynamic behavior they give rise to has a claim to reality, as a new form of "life."

Langton shaped these ideas into the guiding principles of a new research program he dubbed "Artificial Life," which was officially launched with a conference devoted to the "Synthesis and Simulation of Living Systems" he organized at Los Alamos

---

[10] See the essays in Theory and Applications of Cellular Automata, edited Stephen Wolfram (Singapore: World Scientific, 1986).
[11] Christopher G. Langton, "Artificial Life," in Artificial Life, ed. Christopher G. Langton (Reading, Mass.: Addison-Wesley, 1989), p. 41

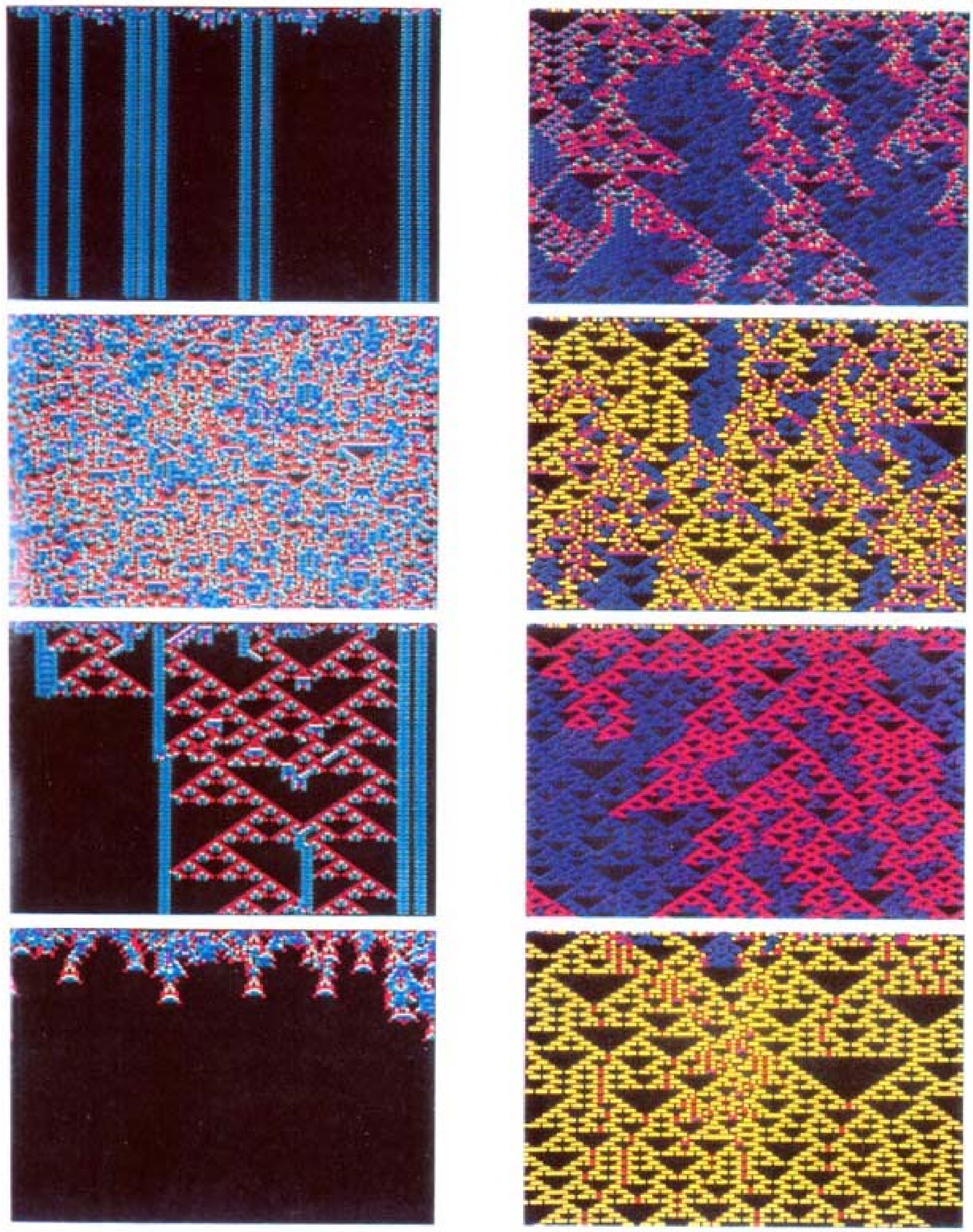
in 1987. The conference drew together a diverse assortment of 160 scientists working on "a wide variety of models of living systems, including mathematical models for the origin of life, self-reproducing automata, computer programs using the mechanisms of Darwinian evolution to produce co-adapted ecosystems, simulations of flocking birds and schooling fish, the growth and development of artificial plants, and much, much more."[13] Appropriately, the title of the published conference proceedings was Artificial Life, and Langton was soon widely known as its founder.[14]

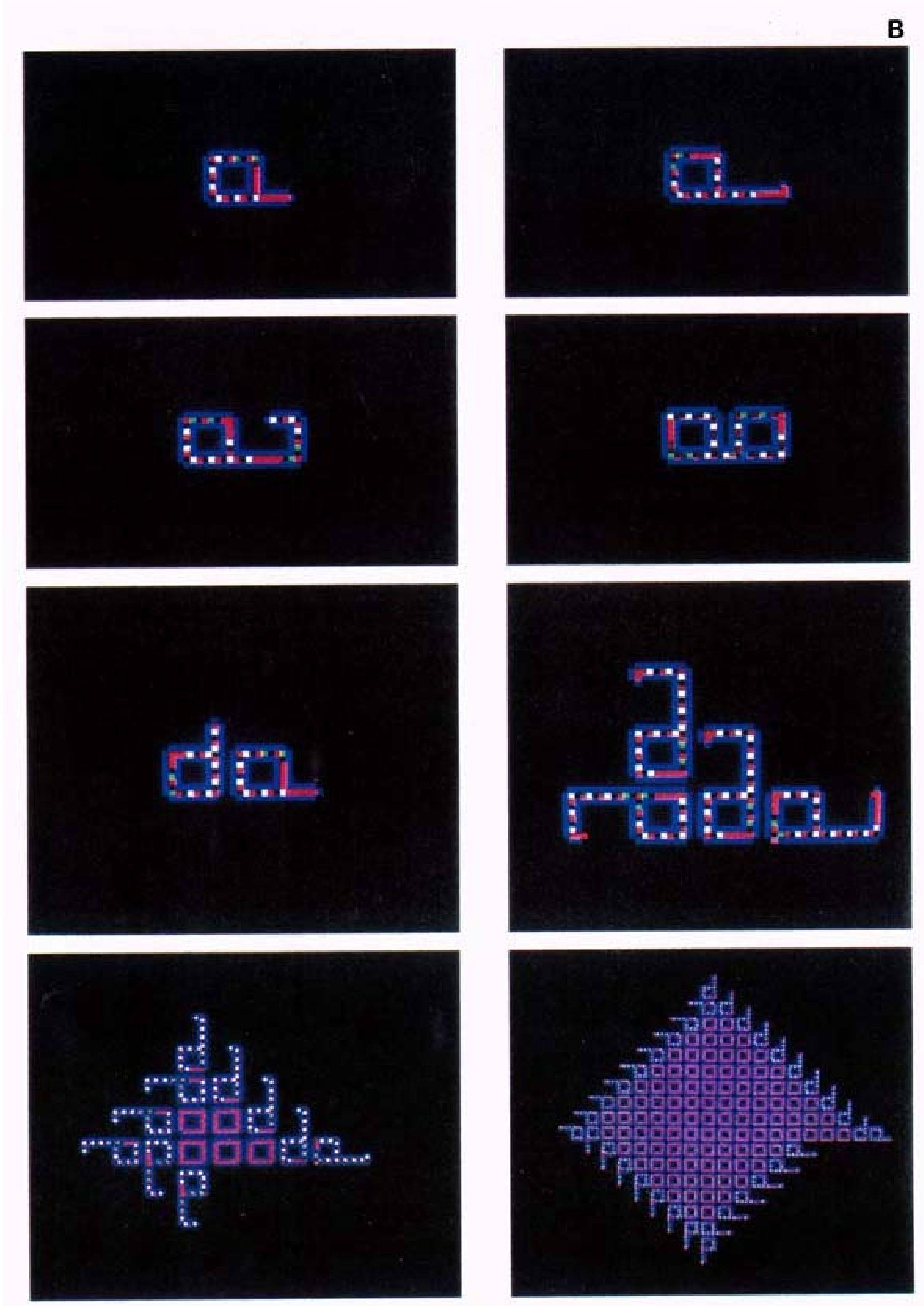
Langton's path to conceptual innovation had begun simply enough, as an attempt to pursue a mysterious feeling he had experienced while working on the mainframe computer at Massachusetts General Hospital, which employed him as a programmer. One night while debugging code he had John Conway's Game of Life running on the computer (often the case with programmers at the time). Suddenly he was overwhelmed by a sense that what was on the screen was somehow "alive," that something akin to life was evolving before his very eyes. His next step was to seek out books on cellular automata theory and then to retrace von Neumann's work. A telephone call to Arthur Burks brought the unexpected news that a self-reproducing cellular automaton had yet to be simulated on a computer. Fortunately for Langton, desktop computers were beginning to appear on the market, and in the summer of 1979 he bought an Apple II. Within months he was able to duplicate Codd's work with eight-state cellular automata, and by October had created little digital loops with short tails -- were they organisms or machines? -- that could reproduce themselves. (Color photographs of cellular automata and Langton's digital organisms are reproduced as A and B below).

---

[12] Langton's idea that life comprehends a virtual dimension because it arises from nonlinear interactions among many physical parts has never been adequately emphasized, at least to my knowledge.
[13] Langton, Artificial Life, Preface, p. xv.
[14] Steven Levy, in Artificial Life: A Report from the Frontier Where Computers Meet Biology (New York: Vintage Books, 1992), calls von Neumann the "father" and Langton the "midwife" (p. 93). Levy's book, the source of the following biographical details, provides the most detailed historical account of the development of Artificial Life up to about 1990.

A





The role of the computer in Langton's research and formulations cannot be overstated. Although Neumann's work on cellular automata theory had followed rather closely upon his work developing the new computing machines (most notably his design for EDVAC, the first electronic stored program computer), the two projects do not appear to have had any necessary or intrinsic connection. For Langton, however, A-Life is inseparable from new ideas about computation. In his essay "Artificial Life," which serves as the introduction to the published conference papers, Langton includes a section entitled "The Role of Computers in Studying Life and Other Complex Systems" in which he makes several key points. [15] Whereas Artificial Intelligence, he says, takes the technology of computers as a model of intelligence (the computational paradigm), Artificial Life attempts "to develop a new computational paradigm based on the natural processes that support living organisms" (50). This significant shift was already implicit in the development of "connectionist" models in Artificial intelligence research, which had been developed from McCulloch and Pitts' work with neural nets but then had been largely abandoned when Marvin Minsky and Seymour Papert demonstrated that such networks could not perform the "XOR" computational function. Their demonstration was refuted in the early 1980s, and neural net research –resurrected under the name of Connectionism—was re-established as a viable alternative for AI research.[16] In their rejection of the digital computer as a model of intelligence, however, Artificial Life and Connectionism have more than an acknowledged conceptual affinity; both envision a new conception of computation and consequently a new computational paradigm.

The new computational paradigm inverts the top-down, centralized command and control structure that dominated early AI approaches to modeling intelligence and that is still deployed in most computer architectures today. In contrast to the top-down

[15] Most of the citations that follow are taken from the revised and extended version reprinted in The Philosophy of Artificial Life, ed. Margaret A. Boden (Oxford: Oxford University Press, 1996). In some instances, however, citations are taken from the original version published in 1989. In those cases that date is also given with the page number.

approach, Artificial Life "starts at the bottom, viewing an organism as a large population of *simple* machines, and works upward *synthetically* from there – constructing large aggregates of simple, rule-governed objects which interact with one another nonlinearly in the support of life-like, global dynamics" (2, 1989). The result is what Langton calls "emergent behavior":

> Natural life emerges out of the organized interactions of a great number of nonliving molecules, with no global controller responsible for the behavior of every part. Rather, every part is a behavior itself, and life is the behavior that emerges from out of all of the local interactions among individual behaviors. It is this bottom-up, distributed, local determination of behavior that AL employs in its primary methodological approach to the generation of life-like behaviors. (2-3, 1989)

It would be inappropriate, however, to instantiate this bottom-up approach using a traditional computer program –"a centralized control structure with global access to a large set of predefined data-structures"—since such programs are specifically designed to produce a final result, not allow "ongoing dynamic behavior." Langton therefore proposes the following essential features of computer-based Artificial Life models:

- They consist of populations of simple programs or specifications.
- There is no single program that directs all of the other programs.
- Each program details the way in which a simple entity reacts to local situations in its environment, including encounters with other entities.
- There are no rules in the system that dictate global behavior.
- Any behavior at levels higher than the individual programs is therefore emergent.

(3-4, 1989)

These criteria, as we'll later see, define a computational paradigm that will exceed its application to Artificial Life, and later be generalized in what is now called Complexity Theory. (Langton himself has contributed to it by developing a software program called "Swarm" used for the simulation of highly distributed, multi-agent systems.)

---

[16] These issues are discussed in detail in Chapter 5 below.

As Langton goes on to point out, the computer is both "an alternative medium within which [one can] attempt to synthesize life" and a laboratory tool that replaces the "wetware paraphernalia" that would normally stock a typical biology lab. This is because the computer itself can incubate informational structures. As Langton puts it: "Computers themselves will not be alive, rather they will support informational structures within which dynamic populations of informational 'molecules' engage in informational 'biochemistry'" (50-51). As both medium and tool, computers can not only simulate complex processes such as turbulent flow but also show that "complex behavior need not have complex roots" (51); indeed, one of Langton's major points is that complexity often arises from the interactions of many simple elements. Summarily, then, two important themes — that complexity is an emergent, bottom-up phenomenon arising from decentralized forms of computation, and that we can think about "life" in terms of an exchange between informational structures— will be fundamental to Artificial Life. And both owe their condition of possibility to the fact of the computer.

Artificial Life: The First Formulations

In his programmatic essay, "Artificial Life," Langton recontextualizes his earlier research with cellular automata, for the most part summarized in articles published in Physica D,[17] within a broad overview of a developing new field of scientific inquiry. Thus, in addition to his own work, he discusses the work of those he deems necessary or exemplary for this new research. But beyond conveying "the 'spirit' of the Artificial Life enterprise" (92), Langton opens up and marks out a beginning that will necessarily come to focus on the conditions of a moment of origination, ostensibly the origins of life but also the origins of a new kind of science grounded in a re-conceptualized understanding of the relationship between computation and life based on their simulation and synthesis. In these terms it is not enough to say that the computer --as both new tool and medium--

makes research in Artificial Life possible. Inasmuch as the computer makes it possible not only to bring into existence non-organic life-like forms but to forge a new conceptual understanding of the mechanisms and even the conditions of possibility of organic life as well, the computer's significance must be considered of ontological import. Moreover, with the advent of Artificial Life "the prodigious idea of *Nonorganic Life*," as Deleuze and Guattari put it, becomes a sanctioned objective of the scientific enterprise itself. In these

terms, Langton and his fellow A-Life scientists must also be considered among the ranks of what we might call, following D&G, "silicon probe heads." And yet, as its brief history has already made clear, Artificial Life is less a "window" onto the machinic phylum than its burgeoning and increasingly rich extension in new forms. With this perspective in mind, I want to review the general intellectual framework that Langton proposes.

In contrast to biology's analytic approach, which breaks down living systems into their component parts in an attempt to understand how they work, Artificial Life seeks to synthesize not simply "life-as-we-know-it" but more importantly "life-as-it-could-be." Theoretically this implies that Artificial Life takes a performative and productive, and no longer a mimetic, approach, but wisely Langton confines himself to what it means in practice: investigating life-like behavior and the appearance of diversity through evolution in an artificial or wholly constructed medium. In its attempts to generate life-like behavior, then, A-Life will first identify the mechanisms by which behavior is generated in natural systems and then recreate these mechanisms in artificial systems. Thus, unlike Artificial Intelligence, which seeks to produce the effects of intelligence without concerning itself with the methods by which it occurs naturally, A- Life endeavors to follow nature in one fundamental aspect, which Langton succinctly emphasizes: "Living systems are highly distributed and quite massively parallel" (41). Furthermore, if life results from a particular organization <u>of</u> matter, rather than from something inherent

---

[17] See Langton, "Self-Reproduction in Cellular Automata," <u>Physica 10D</u> (1984), 135-144; and "Studying

<u>in</u> matter, then nature suggests that this organization emerges from dynamic, nonlinear interactions among many small parts or elements. In other words, life does not arise from the "top-down" imposition or infusion of some universal law or life-principle on lower, more localized levels of activity but emerges spontaneously from the "bottom-up." Having rejected vitalism, modern biologists generally believe that life can be explained by biochemistry. In principle, Langton points out, this means that they believe that "living organisms are nothing more than complex biochemical machines" (5). In his view, however, a living organism is not a <u>single</u>, complicated biochemical machine, but a large <u>population</u> of relatively simple machines:

The complexity of its behavior is due to the highly nonlinear nature of the interactions between all of the members of this polymorphic population. To animate machines, therefore, is not to "bring" life to a machine; rather it is to organize a population of machines in such a way that their interactive dynamics is "alive." (5, 1989)

We can't help but note that from this agenda—to organize a population of machines—it is only a short step to another: to create the conditions in which a population of machines can self-organize.

To model the complexity of behavior characteristic of life as a multiplicity of machines Langton distinguishes between the genotype and the phenotype. The first term refers to the machinery or local rules that produce the local agents or elements of the system, the second to the behavior that results from their interactions. More simply, the genotype is a "bag of instructions," the phenotype what happens when those instructions are executed. These two notions are generalized so that they can be applied to non-biological situations, yielding, as a consequence, a distinction of levels: at the level of what Langton calls the GTYPE local rules produce simple nonlinear interactions, whereas at the level of the PTYPE global behaviors and structures emerge. Thus defined, the model exhibits the essential key features of rich behavior and unpredictability:

Artificial Life with Cellular Automata," <u>Physica 22D</u>, (1986), 120-149.

Nonlinear interactions between the objects specified by the GTYPE provide the basis for an extremely rich variety of possible PTYPES. PTYPES draw on the full combinatorial potential implicit in the set of possible interactions between low-level rules. The other side of the coin, however, is that we cannot predict the PTYPES that will emerge from specific GTYPES, due to the general unpredictability of non-linear systems (57-58).

In other words, one cannot look at a GTYPE and determine what kind of behavior or properties it will generate in the PTYPE. Inversely, one cannot begin with a desired behavior or property (evident in a specific PTYPE) and work backwards to derive the specific GTYPE that produced it. This is because any specific PTYPE is the effect of many nonlinear interactions among local elements. Langton offers the example from biology: What changes would have to be made in the human genome in order to produce six fingers on each hand instead of five? No answer can be calculated; there is only trial and error. Or rather, there is nature's way: trial and error guided by the process of evolution by natural selection. Langton concludes: "It is quite likely that this is the only efficient, general procedure that could find GTYPES with specific PTYPE traits when nonlinear functions are involved" (58). Thus evolution enters the picture not as an external theme that Artificial Life will have to address, but as the inevitable and necessary solution to a dilemma internally generated by the very nature of the model.

When Langton turns the building of actual GTYPE/PTYPE systems, all of the examples he considers are based on the methodology of "recursively generated objects," that is, objects generated by feeding the output of one calculation back into the input for the second, and so on indefinitely. The appeal of this approach, he explains, "arises naturally from the GTYPE/PTYPE distinction: the local developmental rules —the recursive description itself— constitute the GTYPE, and the developing structure —the recursively generated object or behavior itself—constitutes the PTYPE" (59). The resultant behavior, consequently, occurs in the same "medium" as the rules written to generate it. Specific examples are taken from three areas: Lindenmayer systems (or L-systems), cellular automata, and computer animations.

L-systems are produced by simply reiterating a set of substitution rules, as in: for A, substitute CB, for B substitute A, for C substitute DA, and for D substitute C. Taking A as the initial "seed," one can quickly generate a "linear growth" over a relatively short number of succeeding time-steps. Thus A >CB > DAA > CCBCB, etc. This kind of "growth" can be correlated with the behavior of a specific type of finite-state machine. To achieve a "branching growth" one only has to introduce "context sensitive" rules, such that the substitution rule will change depending on what lies to the left or right of the element in question.[18] With context sensitive rules one can generate a grammar similar in type to what Noam Chomsky defines as a "regular language," as well as "propagate a signal" in the sense of moving a symbol from one position to another in a symbol string -- for example, from the far left to the far right positions. Ultimately, context sensitive rules mean that one can embed a computational process that will directly effect the structure's development.

This is made evident in the example of cellular automata, where Langton introduces his own work, specifically the q-shaped loops discussed above as the "simplest known structure that can reproduce itself" (64). In the lattice of finite automata the transition- state tables constitute the local rule set or GTYPE. Basically, through a more extended application of context sensitivity rules, Langton has been able to "embed general purpose computers" (64), making it possible for structures to compute and construct other structures. Moreover, since these computers are "simply particular configurations of states within the lattice of automata, they can compute over the very set of symbols out of which they are constructed" (64, Langton's emphasis). This means that the transition-state tables defining the states of the cellular automata constitute the "physics" of a space/time universe.[19] Within this universe, embedded at a second level

[18] See Langton, pp. 60-61 for concrete illustrations.
[19] In fact, in "Digital Mechanics," Physica D 45 (1990), 245-270, Edward Fredkin has argued that the universe is fundamentally computational, and functions as a giant cellular automaton.

and beginning with a seed structure, recursively applied rules generate a loop that reproduces itself, then each of the two loops in turn reproduce themselves and so on, eventually producing a colony of loops. What is most significant about this "double level of recursively applied rules" is that the system generates itself autonomously. As Langton explains, the resulting system "makes use of the signal propagation capacity to embed a structure that itself computes the resulting structure, rather than having the 'physics' directly responsible for developing the final structure from a passive seed"(65). What is captured, consequently, exhibits the flavor of what happens in natural biological development, where the genotype codes for the constituents of a dynamic process in the cell and this dynamic process then "computes" the expression of the genotype.

For his third example Langton cites Craig Reynolds' computer simulation of flocking behavior. Reynolds discovered that only three simple and easily programmable rules were necessary to make his artificial birds or "Boids" exhibit realistic flocking behavior:

- to maintain a minimum distance from other objects in the environment, including other boids,
- to match velocities with Boids in the neighborhood, and
- to move towards the perceived center of mass of the Boids in its neighborhood.

(Langton, 66)

This vivid example allows Langton to make an essential point regarding the "ontological status of the various levels of behavior in such systems," namely, that even though Boids are obviously not real birds, "flocking Boids and flocking birds are two instances of the same phenomenon: flocking" (68, Langton's emphasis). The flocking behavior of Boids is not only life-like but emerges within an artificial system in the same way that it emerges in nature. What matters is not so much that Reynolds' work enables us to understand the mechanism of flocking (a secondary benefit) as that it demonstrates how complex behavior (the PTYPE) emerges from the interaction of simple agents whose local behavior is determined by simple rules (the GTYPE). This complex behavior, moreover,

is just as real or genuine as its naturally occurring counterpart. Thus, if Langton's argument is valid, we must now revise our terms, since the words "simulation" or "imitation" are no longer appropriate or precise. Reynolds' boids exhibit flocking behavior, *tout court*, not the simulation or imitation of flocking behavior. Langton suggests as much when he refers back to the L-systems and self-reproducing loops in a summary statement: "The constituent parts of the artificial systems are different kinds of things from their natural counterparts, but the emergent behaviors that they support are the same kinds of thing as their natural counterparts: genuine morphogenesis and differentiation for L-systems, and genuine self-reproduction in the case of the loops" (68). Langton concludes this section with a re-statement of "the big claim":

> A properly organized set of artificial primitives carrying out the same functional roles as the biomolecules in natural living systems will support a process that will be 'alive' in the same way that natural organisms are alive. Artificial Life will therefore be <u>genuine</u> life—it will simply be made of different stuff than the life that has evolved here on Earth.
> (69, Langton's emphasis)

Having established the cogency of this fundamental idea, Langton turns to evolution, and to how it can be engineered in Artificial Life research. In his original essay Langton considers only genetic algorithms, developed by John Holland and discussed at length in his <u>Adaptation in Natural and Artificial Systems</u> (1975). Essentially the genetic algorithm is the result of methods Holland invented to 'breed' algorithms that are more efficient at performing specific tasks. You start with a population of algorithms (the algorithm would typically be a simple symbol string), then select the most successful and apply 'genetic operators' which generate a population of offspring. The basic genetic operators are: 1) random mutation 2) inversion, or bit-flipping and 3) "crossover," where the algorithm is split in half and the two halves "mated" with the two halves of another algorithm. The entire process is then repeated several times. The method turns out to be a very robust way to search the "schema space" of possible algorithms, and rapidly produces algorithms meeting ever-higher fitness criteria. Holland also produced several theorems explaining this remarkable efficiency.

In the expanded version of his essay Langton mentions several more significant research efforts (which will be discussed in more detail in Chapter 4). In connection with Holland's work he mentions John Koza's development of techniques for genetic programming. As an example of "computational artificial selection," he discusses Richard Dawkins' "biomorph" breeder program that Dawkins wrote to illustrate the ideas in his book, The Blind Watchmaker. Evolutionary programming techniques or "algorithmic breeders," as Langton call them, are also used by Danny Hillis to design more optimized sorting programs and by Kristiann Lindgren to evolve better strategies for playing Prisoner's Dilemma. Throughout this section Langton considers the various means that have allowed researchers to progress toward the point where the human hand is eliminated from the selection/breeding process and something like genuine "natural" evolution can occur in a computer. In these terms his crowning example is Thomas Ray's computer-created artificial world Tierra, which represents the fullest and most successful instantiation of the dynamic of evolution in Artificial Life. In Tierra, digital organisms (actually blocks of code) reproduce, mutate and compete for the computer's resources, with the result that a complex ecology rapidly evolves. However, since this work will be discussed in the next chapter, I now turn to Langton's more recent research on computation and the origin of life, as set forth in his essay "Life at the Edge of Chaos."

The Origin of Life as the Origin of Information

Whereas in "Artificial Life" Langton sets out a new framework for programmatic research on the "biology of possible life," that is, life as it may be defined without the historical and possibly contingent restrictions of carbon-based life as it has developed on the planet earth, in "Life at the Edge of Chaos" he attempts to answer a very specific question: "under what conditions can we expect a dynamics of information to emerge spontaneously and come to dominate the behavior of a physical system?" (42).[20] The question, of course, directly assumes that what characterizes life or living systems is that there "information processing has gained control over the dynamics of energy, which determines the behavior of most non-living systems" (41). Langton's answer, simply enough, is emblazoned in his title: it is "at the edge of chaos" that information-processing will gain the upper hand over the dynamics of energy exchange. One could, to be sure, quarrel with the background assumption. One could ask, for example, if the Stock Market is a living system, and if it is always so easy to draw a clear line between information-processing and the dynamics of energy exchange. But Langton himself is not concerned with such questions, and proceeds directly to what will be the burden of his essay: to provide substantial argumentation and convincing evidence for his claim.

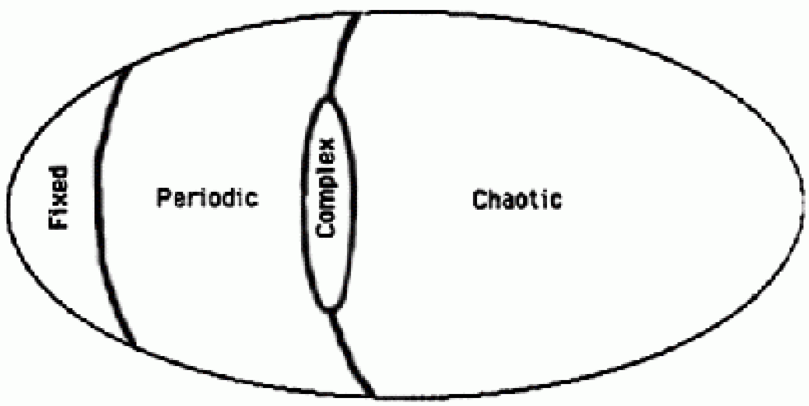
The evidence, once again, will come from his work with cellular automata. He justifies the appropriateness of CA according to the following criteria:

1) As nonlinear dynamical systems, CA exhibit the entire spectrum of dynamical behaviors, from fixed-points, through limit cycles, to fully developed chaos.
2) CA are capable of supporting universal computation. Thus they are capable of supporting the most complex known class of information dynamics.
3) There is a very general and universal representation scheme for all possible CA: a look-up table. This form of representation allows us to parameterize the space of possible CA, and to search this space in a canonical fashion.
4) CA are very physical, a kind of "programmable matter." Thus, what we learn about information dynamics in CA is likely to tell us something about information dynamics in the physical world. (42)

The very terms here --that CA are a kind of “programmable matter,” exhibiting the entire spectrum of dynamical behavior-- again suggest that Langton’s computer experiments are directly related to D&G’s machinic phylum. But of course this is not his framework; what he seeks, basically, is to demonstrate a correlation between the information-processing capacity of the CA and their dynamical state or regime. In order to do this, he devises a "tuning knob" or “lambda parameter,” as he calls it, that reflects how changes

[20] Published in Artificial Life II, ed. Christopher G.Langton *et al* (Redwood City, CA: Addison-Wesley, 1992), 41-91. Page numbers of citations will be inserted directly into the text.

in CA rules correspond to the changing behavior of the cellular automata, as this behavior is mapped in a phase space in exactly the same way that the behavior of dynamical systems is mapped in chaos theory.[21] The lambda parameter varies over a spectrum of values ranging from a lifeless or frozen state of the automata (given by 0) to absolute chaos (given by 1). At low values near zero the system is rigid and exhibits little life; its behavior thus seems to be determined by a fixed point or periodic attractor. For high values close to 1, on the other hand, the system goes chaotic, as if pulled by a strange attractor. The area of greatest interest lies somewhere in between, in "the sweet spot" of the CA rule space that Langton indicates with this diagram:



Schematic drawing of CA rule space indicating relative location of periodic, chaotic, and complex regimes.

As the behavior of the CA moves from fixed or periodic to chaotic, Langton argues, it does not simply jump from one regime to another but passes through the space of a critical phase transition. This transition regime, however, "is not simply a smooth surface between the other two domains" but a complex domain with its own "complicated structure" (75), meaning that it can support "complex interactions between propagating and static structures" (69). But if this is the case, these structures in turn

[21] For more on these concepts and nomenclature (portraits in phase space, attractors, etc.), see Ralph H. Abraham and Christopher D. Shaw, Dynamics: The Geometry of Behavior (Redwood City, CA: Addison-Wesley, 1992).

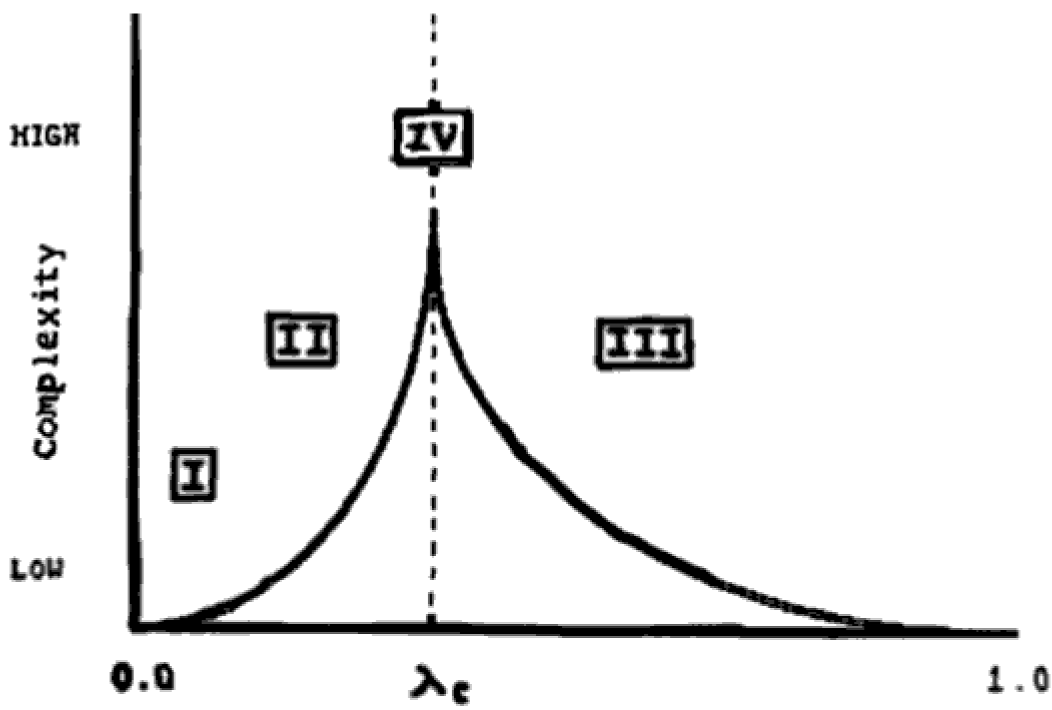
"can be pressed into useful service as logical building blocks in the construction of a universal computing device" (69). Langton offers a demonstration of how this might work: glider guns from Conway's Game of Life are constructed to emit on-off pulses that can be arrayed together to function as AND, OR and NOT logic gates, which are then used as the basic building blocks of universal computation. This array of glider guns instantiates what Langton calls "constructability" -- that is, it is one way that complex CA rules in this phase transition space can enable universal computation to occur.

However, as Langton acknowledges, there is no absolute guarantee or sure determination: "similar constructions can be made for other complex rules, but not all complex rules tried have yielded to such simple constructions" (75). Yet this is not a stumbling block, since what matters is the remarkable similarity observed between "the surface-level phenomenology of CA systems" and "the surface-level phenomenology of computational systems," specifically in respect to "the existence of complexity classes, the capacity for universal computation, undecidability, etc." (75). For Langton, this fundamental similarity can only be explained by concluding that "the structure of the space of computations is dominated by a second-order phase transition" (75). In other words, the complex behavior of CA and the conditions of universal computation exhibit the same underlying structure because both are instances of a more general structure of change, which can be characterized as a second-order phase transition. Let us therefore look more closely at this general underlying structure.

Thanks to Stephen Wolfram's work on cellular automata, Langton knew that their behavior falls into four general classes. In Class I, the cells go quiescent or "die" within several time-steps. (In dynamical systems theory, they are drawn to a single-point attractor.) In class II, the cells are livelier, but soon settle down to static or periodically oscillating configurations. (They are drawn by a periodic attractor). In Class III, the cells behave chaotically, with no patterns forming or lasting for more than a time-step or two.

(They are drawn to a strange attractor.) In Class IV, where the behavior is the most interesting, the cells never settle down but endlessly form patterns, break apart, and reform into other patterns. (This is the kind of complex behavior often exhibited in Conway's Game of Life.) Neither periodic nor chaotic, the behavior of the cells in Class IV corresponds to nothing clearly understood in dynamical systems theory.

Nevertheless, as Langton observes, the behavior of Class IV CA corresponds exactly with the lambda values that mapped the space of a phase transition in the CA rule space diagram. Whereas low lambda values correspond with Classes I and II, and high values with Class III, the complexity of Class IV behavior seems to correspond to this peculiar space, as if it defined a specific regime with its own defining characteristics. (See Langton's diagram below.) Approaching this regime from low lambda values, the CA would exhibit longer and longer periods of "transient time," while showing more and more sensitivity to array size. When moving away from it toward high values the exact inverse would occur. At values that fell exactly within the domain of complexity the transient time would become unmeasureably long or undecidable. Langton thus concluded that the CA phenomenology was structured by an underlying phase transition.



Schematic drawing of complexity versus λ over CA space, showing the relationship between the Wolfram classes and the underlying phase-transition structure.

Langton also concluded that the same or a very similar phenomenology held true for the space of computation. Most of the time mathematicians have a practical sense of whether particular computations can be carried out in finite time or not. Alan Turing showed however that there are many instances where computability cannot be determined in advance. Problems, therefore, can be classified accordingly: the first class is comprised of computations that "halt" (the computation is solved) in finite computational time; the second of those that are "non-halting" (and thus that cannot be computed); and the third of those that are undecidable. Not surprisingly, the phenomenology of these three classes suggested to Langton the same structure of the critical transition regime: the halting computations correspond to Class I and Class II CA, non-halting computations to Class III CA, and undecidable computations to Class IV. Once again, this would mean that as one approached the complex domain from either side the transient time of computability would increase exponentially -- in short, that the most complex computations are to be found at a phase transition.

Summarily, then, Langton was seeking to establish a clear correlation between the complex behavior of nonlinear dynamical systems (in this instance, cellular automata) and the structure or conditions of "computability." Hence the question: What kind of a structure is necessary if information is to be computed? The phase transition regime turns out to provide the underlying structure and the key to understanding both phenomena. To clinch his argument Langton presents a third instantiation: the phase transitions of matter itself. We are all familiar with its three states: solid, liquid and gas. As ice melts, for example, it moves through a phase transition from solid to liquid in which each molecule has to "decide" between two qualitatively different states. Moreover, as the temperature approaches the limits of the phase transition, the molecules require more and more time to reach a decision (and thus their "transient time" is said to increase). If Langton is right, the phase transition is thus a moment of critical "slowing down," when "the system is engaged in 'solving' an intractable problem"

(82). Langton acknowledges that this sounds anthropomorphic, but he also feels that the results of his work justify our thinking that such systems are "effectively *computing* their way to a minimum energy state" (82, Langton's emphasis).

What all this points to is "evidence for a *natural domain of information* in the physical world" (81, Langton's emphasis). Needless to say, whether information exists in the natural world or is only a human conceptualization is hardly a neutral question; it raises not only the semantic problem of how information is to be defined but also the thorny epistemological issue of its status. For Langton, however, the evidence that information-processing --the storage, transmission and modification of information-- takes place in the physical world is not in question.[22] Nor is there any question about where we should expect this process to emerge and come to dominate the dynamics of a physical system: it is in the vicinity of a critical phase transition, that is, in a space between orderly and chaotic regimes, and it can be defined mathematically using the lambda parameter.

Having therefore answered the question posed at the outset, in the essay's final section Langton summarizes in more speculative terms what this work implies about the relationship of life to matter, computation (or intelligence), and evolution. It seems very likely, he suggests, that "the origin of life occurred when some physico-chemical process underwent a critical phase transition in the early history of the earth" (81). His own work lends credence to this claim by offering compelling (if not exhaustive and incontrovertible) evidence for a series of subsidiary claims. First, there is "a fundamental equivalence between the dynamics of phase transitions and the dynamics of information processing" (82). Not only does the phenomenology of phase transitions explain the phenomenology of computation, but also the reverse; in fact, these are not

---

[22] See the special issue of Physica D 42 (1990) organized by Stephanie Forrest devoted to "emergent computation." Langton's contribution, "Computation at the Edge of Chaos: Phase Transitions and Emergent Computation," 12-37, treats the informational process more explicitly.

two phenomenologies but one: "We are observing one and the same phenomenon reflected in two very different classes of systems and their associated bodies of theory" (82). Second, solids and liquids are dynamical, rather than material categories, and these two universality classes of dynamical behavior are separated by a phase transition.[23] Since the dynamics of systems operating near this phase transition provide a basis for embedded computation, "a third category of dynamical behaviors exists in which materials—or more broadly, material systems in general—can make use of an intrinsic computational capacity to avoid either of the two primary categories of dynamical behaviors by maintaining themselves on indefinitely extended transients" (83). More simply, a dynamical system operating near a phase transition can use its acquired computational capacity to maintain itself near this enabling regime. Third, "living systems can perhaps be characterized as systems that dynamically *avoid* attractors," that "have learned to steer a delicate course between too much order and too much chaos—the *Scylla* and *Charybdis* of dynamical systems" (85). In these terms evolution can be viewed "as the process of gaining control over more and more 'parameters' affecting a system's relationship to the vital phase transition" (85). In sum, after a living system has emerged near a critical phase transition, "evolution seems to have discovered the natural information-processing capacity inherent in near-critical dynamics, and to have taken advantage of it to further the ability of such systems to maintain themselves on essentially open-ended transients" (85).[24]

Having started with the theoretical objective to synthesize "artificial life" by abstracting and simulating certain of its complex and dynamical behaviors, Langton here arrives at a scenario for understanding not only the basic conditions of life but a specific

[23] Langton explains the elimination of gases from consideration as follows: "As it is possible to continuously transform liquids into gases and vice-versa without passing through a phase transition, they are taken to constitute a single, more general phase of matter: fluids" (83).
[24] As we'll see in Chapter 4, biologist Stuart Kauffman essentially agrees, having found supporting evidence in his own work with gene networks and autocatalytic sets. Simply put, Kauffman believes that life flourishes at the edge of chaos because evolution takes it there, in order to produce further evolution.

mechanism by which life might have arisen from non-life, and then maintained and perpetuated itself. Having argued along the way that "hardness, wetness, or gaseousness are properties of the organization of matter, rather than properties of the matter itself" (84), thereby implying that it is only a matter of organization to turn "hardware" into "wetware" and that hardware should eventually be able to achieve everything achieved by wetware, Langton ends with a paean to water that the modernist writer James Joyce would no doubt have appreciated. While theorists of the origin of life have looked to "the dynamics of molecules *embedded* in liquid water," Langton says, it may well be that life "originated in the dynamics of water itself" (85).

Autopoiesis

The sheer intellectual interest and importance of Langton's work should be self-evident, the hostile silence of certain traditional-minded biologists notwithstanding. Like Robert Shaw in his work on chaos, Langton in his foundational work on A-Life has combined and synthesized a computational, information-processing approach with a dynamical systems perspective, making the quantitative and qualitative two aspects of the same discursive assemblage. Later I will consider Melanie Mitchell and James Crutchfield's "friendly" critique of Langton's work on computation at the edge of chaos (one problem is that Langton doesn't define computation in relation to specific tasks), as well as further developments in A-Life that similarly contribute to the fields of theoretical biology and nonlinear dynamical systems theory.[25] But here I would like to turn to another "non-standard" theoretical approach to the definition of life, most readily recognized by its central concept of "autopoiesis." This approach was first developed in the 1970s by the biologists Humberto Maturana and Francisco Varela, who thereby inaugurated what is sometimes referred to as the Santiago or Chilean school. For Maturan and Varela an autopoietic system (and all living systems are autopoietic) is organized in such a way that its only goal is to produce and maintain itself. Directly

[25] See Chapter 4 below; in Chapter 6 I also sketch the influence of A-life on the new AI and robotics.

opposed to the computational and informational approach to the study of life, they also decry the centrality of DNA and genetic coding in modern biology, and seek to put the autonomy and individuality of all living systems on a firm theoretical footing.

In Principles of Biological Autonomy, first published in 1979, Varela re-presented these ideas in a more expanded context with several specific examples.[26] (As we'll see in a moment, Maturana and Varela's first presentation is unremittingly abstract.) More important here, in this later volume Varela discusses his early work with cellular automata, which he used to model a simple autopoietic system that forms self-enclosing boundaries and repairs them when they break down.[27] In the 1990s European scientists found in these formulations the basis for an alternative approach to Artificial Life, no doubt spurred by the first European conference on Artificial Life in 1991, which Varela organized with Paul Bourgine. (The conference papers were published in 1992 in an influential volume entitled Toward a Practice of Autonomous Systems.) What makes this body of work especially significant is less the alternative approach to Artificial Life it offers (in fact, its influence on the "new AI and robotics" is much greater), than the fact that it confirms one of its most fundamental assumptions. Although opposed to the informational/computational basis of Artificial Life, Maturana and Varela share with it a conceptual underpinning that understands living systems as machines.[28]

Before these differences are examined, it should be noted that Maturana and Varela's intellectual lineage also reaches back to cybernetics, or, more accurately, to its "second wave" inaugurated by the work of Heinz von Foerster. Varela himself attests to von Foerster's importance for both the development of cybernetic theory and Maturana's

---

[26] This is particularly true of the revised and expanded French version, Autonomie et Connaissance: Essai sur le vivant (Paris: Editions du Seuil, 1989).

[27] Having picked up the idea from von Neumann, Varela worked with cellular automata in the early 1970s, long before "the Artificial Life wave hit the beach," as he himself puts it in the "The Emergent Self," in The Third Culture, ed. John Brockman (New York: Touchstone Books, 1995), p. 211.

[28] Of course this is true for many biologists; Richard Dawkins for example, as we saw in the Introduction, defines life exclusively in terms of information machines.

and his own work.[29] Having been an active participant and editor of the transactions of the Macy Conferences, von Foerster went on to do original research as the director of the Biological Computer Laboratory at the University of Illinois at Champaign-Urbana from 1960 to1975. Two key ideas he developed there are especially noteworthy: first, that in complex systems “noise” can bring about a higher level of organization, the “order-from-noise principle” he called it, which would eventually lead to theories of self-organization; and second, that self-reference and hence the role of the observer play a constitutive role in the formation of systems.[30] This work, in effect, made von Foerster the architect of second-order cybernetics. In fact, one could easily pursue the claim --made by Varela and others—that von Foerster’s work was unjustly overshadowed and even repressed because of the importance "granted" (in every sense) to dogmatic forms of cognitive science and classical Artificial Intelligence, particularly to what later became known as symbolic or "high church computationalism" (Daniel Dennett's phrase). But for my purposes here what is most important is that von Foerster’s first idea became crucial for complexity theory and his second for Maturana and Varela's theory of autopoiesis, to which I now turn.

Maturana and Varela first developed the idea that living systems are a special type of machine in their collaborative book, <u>Autopoiesis: The Organization of the Living</u>. Published in Chile in 1972, it appeared in English in 1980, together with Maturana's earlier book, <u>The Biology of Cognition</u>, under the collective title, <u>Autopoiesis and Cognition: The Realization of the Living</u>. No doubt one of the most abstract books on biology ever written, it attempts to re-define its fundamental principles in terms of a science of living systems. In so doing it draws tacitly on systems theory (G. Spencer-Brown) and second-order cybernetics (Heinz von Foerster), as well as Maturana's earlier

[29] See Francisco Varela, "Heinz von Foerster, the scientist, the man," SEHR, vol.4, no. 2 (1995).

empirical studies of vision in frogs (and later pigeons), research which stemmed directly from McCulloch and Pitts' "neural net" theory.[31] Altogether, these influences and research led Maturana to reject the assumed objectivity of science and to formulate his own epistemology.

In the introductory essay Maturana wrote for Autopoiesis and Cognition he recounts how the results of his experimental research on vision actually worked against the assumption of realism and epistemological objectivity that framed it, and thus how this re-orientation necessarily imposed itself.[32] Specifically, Maturana and his group were unable to map the activity of cells in the frog's retina directly onto the contours and colors of the visual world, and were forced to reverse their assumptions and the questions they were asking. As Maturana puts it: "What if, instead of attempting to correlate the activity of the retina with the physical stimuli external to the organism, we did otherwise, and tried to correlate the activity of the retina with the color experience of the subject?" (xv). Pushing further, he was forced to assume that "the activity of the nervous system [is] determined by the nervous system itself, and not by the external world; thus the external world would have only a triggering role in the release of the internally-determined activity of the nervous system" (xv). This perspective led Maturana to treat the nervous system as a system closed on itself, and the "report of the color experience as if it represented the state of the nervous system as a whole" (xv). In these terms the nervous system is both autonomous (“operationally closed”) and coupled to the environment (“interactionally open”).

---

[30] See Cahiers du CREA No 8 (November 1985), which is devoted to "Genealogies of Self-Organization," as well as von Foerster's "On Self-organizing Systems and Their Environment," in Observing Systems (Salinas, CA: Intersystems Publications, 1981). The essay was first published in 1960.

[31] This research began in collaboration with J.Y. Lettvin, W.S. McCulloch, and W.H. Pitts. Their essay, "What the Frog's Eye Tells the Frog's Brain," is re-published in McCulloch's Embodiments of Mind.

[32] Humberto R. Maturana and Francisco J. Varela, Autopoiesis and Cognition: The Realization of the Living (Dordrecht, Holland: D.Reidel Publishing Company, 1980), xi-xxx. Page numbers for citations will be inserted directly into the text.

It follows that perception of distinct objects in the external world does not at all correspond to the stimulation of specific cells in the retina; there is no point-to-point correspondence. Rather, the inflection of the visual field by the appearance of a distinct object is the result of a triggering affect that perturbs the visual system as a whole, which then re-establishes its own equilibrium. In the frog's visual system only certain kinds of perturbation are possible, and these correspond to what the frog can "see" — basically, small objects moving fast-- with little ability to discern large objects moving slowly. To say therefore that the frog's vision is perfectly adapted to its environment, enabling it to catch flies and avoid predatory birds, is true but misses the essential point: that the frog does not so much “see” the world as respond to and interact with selected aspects of it. More precisely, seeing is a perceptual/cognitive linkage with stimuli that have no objective existence outside the activities of the perceiving subject. It is only the observer who infers the distinction, and who may then describe the interaction. However, purportedly scientific descriptions usually involve the attribution of causal relations, and the whole reified metaphysic of realism and objectivity soon follows. In The Biology of Cognition Maturana moves against the current of these dominant assumptions and seeks to replace them with a conceptual apparatus based on "circular organization" and "self-referential systems."

In Autopoiesis: The Organization of the Living, Maturana and co-author Francisco Varela (his former student) deploy these terms in an attempt to answer the age-old question: What is life, and how should it be defined? Their assumption --and point of departure-- is that "there is an organization that is common to all living systems, whichever [*sic*] the nature of their components" (76). As scientists, they openly declare that their explanation of life is mechanistic, not vitalist; yet they also insist that it is in the circular organization of physico-chemical processes, not in the specific properties of the latter, that life is to be found. Focusing on the fact that it is the organization of living systems that enables them to maintain their own boundaries and to replenish their component parts, and hence to maintain their identity over time, Maturana and Varela relegate the specific details of these processes to a secondary concern. This includes what for most biologists are the essential processes of reproduction and evolution. In fact, Autopoiesis: The Organization of the Living evinces little interest in the concrete ‘stuff' of most biological investigations -- the myriad diversity and dynamic profusion of exchanges that take place within and among living entities; instead, like a treatise by Spinoza or Leibniz, it proceeds almost exclusively by definition and conceptual re-framing. What it offers, nonetheless, is a completely new perspective on "life," and how living systems may be thought of as machines.

A machine is an organized unity of various component parts and processes. For Maturana and Varela, the organization of the machine is precisely what gives it this unity; it is what determines "the dynamics of interactions and transformations which it may undergo as such a unity" (77). The structure of the machine, in contrast, is constituted by "the actual relations which hold among the components which integrate a concrete machine in a given space" (77). These definitions allow Maturana and Varela to distinguish sharply between the relations that give the machine its unity and the properties of the components that realize the machine as a concrete system. From this distinction it follows that the organization of a machine is independent of the properties of its components, and thus that a given machine can be realized in different ways and by different components. (In this sense “organization” plays the same role as information in computation: it signifies a functional process or effect not dependent on the exact nature of its supporting material substrate.) A second corollary of these definitions is that the use to which a machine may be put is not a feature of its organization, but rather "of the domain in which the machine operates" (77). Of course we usually think of human-made machines as constructed for a specific purpose or end, including our own amusement. However, just because this aim or purpose is expressed in the product or result of the machine's operation should not lead us to believe that its purpose, aim or function are constitutive properties; these notions, rather, are extrinsic to the machine's organization, and pertain only to the domain of its observation. They may help us to imagine, describe or simply talk about machines, and may even be realized in a particular machine's operation. Nevertheless, they remain in the domain of descriptions generated by the observer.

That living systems are a specific kind of machine therefore cannot be demonstrated by pointing to their component parts or structure, but only to their organization. Specifically, living systems are organized in such a way that they produce and maintain their own identity, an identity that is distinct from the environment and independent of interactions with an observer. Maturana and Varela call this kind of machine "autopoietic" :

> ...an autopoietic machine is a machine organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces the components which: (i) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and (ii) constitute it (the machine) as a concrete unity in the space in which they (the components) exist by specifying the topological domain of its realization as a network. (78-79)

As in the definition of organization above, the emphasis falls on the network of processes that produce and maintain identity, not on the properties of the individual components and their varied relations. It is intentionally a circular definition: A > B > A. Hence, a living entity is a network of processes organized in such a way as to maintain the integrity and functioning of the processes that define it.

As long as a domestic cat (my example) breathes air, drinks water, and eats food, autopoietic networks will provide the energy to generate and maintain the cells and tissue that will enable the cat to interact with its surrounding environment as a cat. In contrast, an automobile is organized in such a way that gasoline is converted into enough kinetic energy to move the automobile across the landscape, but "these processes are not processes of production of components which specify the car as a unity since the components of a car are produced by other processes which are independent of the organization of the car and its operation" (79). Simply put, the car itself does not and cannot maintain its own identity. In Maturana and Varela's terminology, it is therefore an "allopoietic machine." Whereas autopoietic machines are autonomous and subordinate all aspects of their functioning to the maintenance of their own organization (and hence to their identity), an allopoietic machine, as in the example of the automobile, has something different from itself as the product of its functioning. It follows that an allopoietic machine has no individuality and that its identity depends entirely on the observer, the "other" who stands outside its process of operation. This identity is not determined by or through the machine's operation, precisely because the result of this operation is different from the machine itself. An automobile, for example, could serve as a simple vehicle for transportation, a source of spare parts in a junkyard, a collector's item or a fetishized prop in a movie. In contrast, an autopoietic machine maintains its own individual identity independently of its interactions with any observer.

While at first the opposition between autopoietic and allopoietic may seem to re-inscribe the opposition between a living entity and a tool, the organic and the inorganic, Maturana and Varela gradually show that the difference between them cannot be so reduced. Although autopoietic machines are homeostatic, since they maintain as constant precisely the relations that define them as autopoietic, they do not have inputs and outputs, as do allopoietic machines. To be sure, autopoietic machines can be perturbed by independent events and can undergo internal structural changes that compensate for these perturbations. However, adjustments to perturbations, whether as singular or repeated events, are always subordinated to the maintenance of the organization that defines the machine as autopoietic. Nevertheless, human observers can "describe physical autopoietic machines, and also manipulate them, as parts of a larger system that defines the independent events which perturb them" (82). The observer, furthermore, can view these perturbing independent events as input, and the changes the machine makes to compensate for these perturbations as output. As Maturana and Varela often insist, this is precisely the mistake of molecular biology, which treats the living system as an information-processing device. In their terms, this is to treat an autopoietic machine as an allopoietic one. It is also possible "to recognize that if the independent perturbing events are regular in their nature and occurrence, an autopoietic machine can in fact be integrated into a larger system as a component allopoietic machine, without any alteration in its autopoietic organization" (82). In the same way, parts of autopoietic machines can be analyzed as allopoietic "sub-machines" in terms of their input and output. In neither case, however, does the analysis reveal the essential and defining nature of the autopoietic machine.

Having staked out these definitions, Maturana and Varela advance their central claim that "autopoiesis is necessary and sufficient to characterize the organization of living systems" (82). Before attempting to substantiate this claim, they add two points. First, they argue briefly that since living systems are machines, once their organization is understood there is no *a priori* reason why they cannot be reproduced and even designed by humans. To think otherwise would be to succumb to the "intimate fear" that the awe with which we view life would disappear if it were recreated, or to the prejudiced belief that life will always remain inaccessible to our understanding. Second, they point out that as long as the nature of the living organization remains unknown, it is not possible "to recognize when one has at hand, either as a concrete synthetic system or as a description, a system that exhibits it" (83). In other words, it is not always or immediately obvious what is living and what is not. For many biologists, reproduction and evolution appear as constitutive, determinant properties, to which "the condition of living" is subordinated; but as Maturana and Varela point out, once these properties are reproduced in human-made systems those who do not accept that any synthetic or human-made system can be living simply add new requirements.

In the remainder of the book Maturana and Varela develop the implications of their claim that autopoiesis is "necessary and sufficient to characterize the organization of living systems" (82). They argue, for example, against the common assumption that teleonomy or purpose is a necessary feature of a living system. It must be remarked, however, that some of the most interesting implications —that in certain respects, social systems and certain information systems might qualify under their definition as living systems—are not pursued. In effect, Félix Guattari makes this very point in his essay, "Machinic Heterogenesis." [33] Although institutions and technical machines appear to be allopoietic, Guattari notes, "when one considers them in the context of machinic assemblages they constitute with human beings, they become ipso facto autopoietic." Guattari thus finds Varela's concept useful, but only if viewed from the perspective of "the ontogenesis and phylogenesis proper to the mechanosphere superposed on the biosphere" -- in other words, from the perspective introduced by the assemblage itself, which could be said to combine both allopoietic and autopoietic functions. Moreover, Guattari's reflections on certain inherent limits of the concept of autopoiesis lead him to call for a reframing or recontextualization in terms that converge precisely with issues that are central to Artificial Life research: "Autopoiesis deserves to be rethought in terms of evolutionary, collective entities, which maintain diverse types of relations of alterity, rather than being implacably closed in on themselves."

As we'll later see, Varela himself is eventually forced to confront these issues more fully when he considers the immune system. Meanwhile, in Principles of Biological Autonomy he offers a less abstract and more useful summary of autopoietic theory, which he applies to specific biological phenomena. Following the methodological approach Maturana developed in his early theory of vision, Varela considers both the immune system and the nervous system as instances of "operational closure" and

[33] The essay appears as Chapter 2 in his book Chaosmos (Paris: Galilee, 1992). Quotations that follow are taken from Chaosmosis, the translation by Paul Bains and Julian Pefanis (Bloomington: University of Indiana Press, 1995), pp. 39-40.

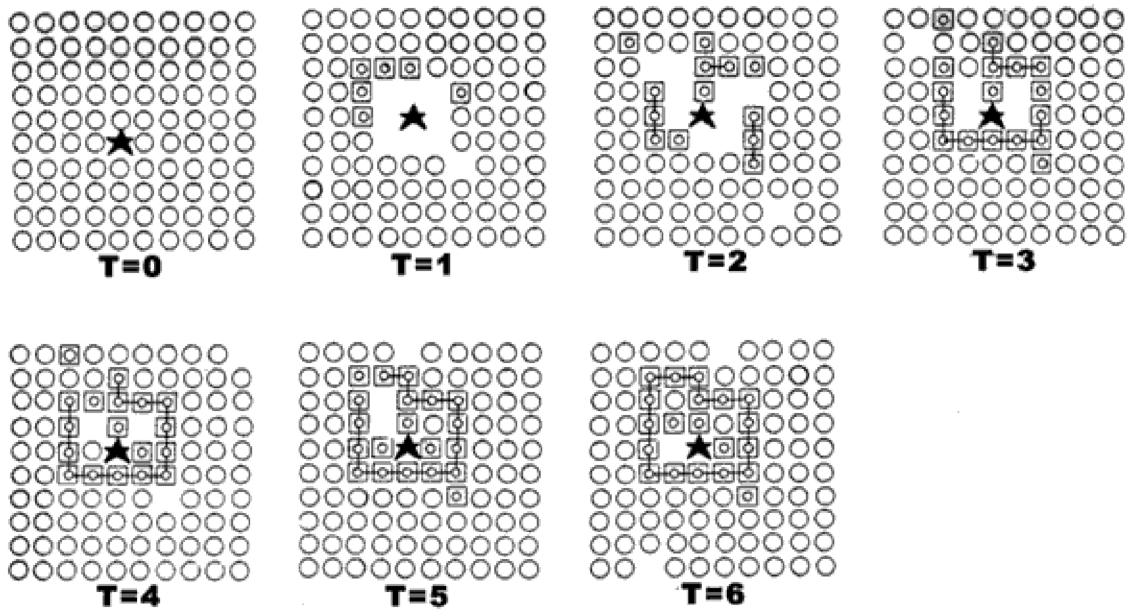
"structural coupling," the technical terms he and Maturana use to describe how a system that works by closing in on itself can nevertheless interact with environmental stimuli.[34] Of more importance here however is the cellular automata model of autopoiesis that Varela includes. Since this model anticipates Varela's explicit interest in Artificial Life, which he prefers to consider under the alternative rubric of "autonomous systems," it is worth considering as a means of contrasting the two approaches.

Toward Autonomous Systems

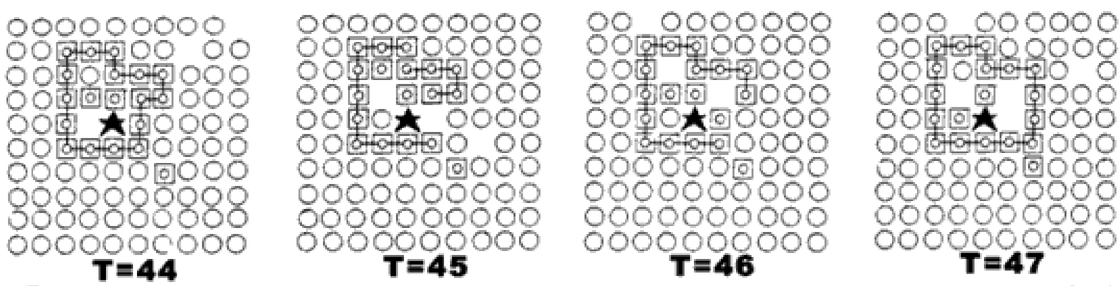
Varela's cellular automaton model demonstrates how an autopoietic unity can spontaneously emerge through a simple linking of elements in the presence of a catalyst. It involves three processes: (1) composition, when two basic elements ("O" and "[ ]") in the presence of a catalyst ("*") form a link (" [O] "); (2) concatenation (or bonding), when a link joins with two or more other links (" [O]-[O] "); and (3) disintegration, when a link decomposes back into two basic elements. Varela had these processes coded in a computer program in such a way as to define the interactions of cellular automata (each space being either empty or occupied by a single element). When the program is run, links form and decompose randomly, but inevitably links will form around a catalyst, and gradually form a boundary in a self-enclosed space. Links will inevitably decompose, but new links will form to replace them, and the boundary will be maintained. This is all illustrated by Varela's diagram:

---

[34] For further examples, see Maturana and Varela's The Tree of Knowledge (Boston: Shambhala Publications, 1987).



A The first seven instants (0→6) of one computer run, showing the spontaneous generation of an autopoietic unity. Interactions between substrate ○ and catalyst ⋆ produce chains of bonded links ⊡, which eventually enclose the catalyst, thus closing a network of interactions which constitutes an autopoietic unity within this universe.



B Four successive instants (44→47) along the same computer run (Fig. 1), showing compensation in the boundary broken by spontaneous decay of links. Ongoing production of links re-establishes the unity under changes of form and turnover of components.

Some years later, Barry McMullin at the Santa Fe Institute pointed out that the coding for the computer program in this model was flawed. In 1996-97, the model was run again by Varela and McMullin, this time with slightly different algorithms to implement the "artificial chemistry" governing the interactions of the elements, and with a different operating system, the "Swarm system" that Langton and his associates had written.[35] Since the results were essentially the same, the updated demonstration further

[35] See Barry McMullin and Francisco J. Varela, "Rediscovering Computational Autopoiesis," in Fourth European Conference on Artificial Life, ed. Phil Husbands and Inman Harvey (Cambridge, Mass. : The MIT Press, 1997), 38-47.

strengthened the claim that "autopoietic phenomena are not dependent on any particular details of the original program or algorithm, but may be expected in any system sharing the same qualitative chemistry" (39, authors' emphasis). The demonstration also brought Varela into direct personal contact with scientists at the Santa Fe Institute, and thus with proponents of what some Europeans thought of as the American brand of Artificial Life. In fact, the term "computational autopoiesis" used in the title of McMullin and Varela's conference paper suggests some sort of *rapprochment*. In fact, however, the term is misleading, since it elides several crucial differences. We can get at these differences by considering some basic questions.

First, how does the autopoietic model and the demonstration of its viability using cellular automata contrast with Langton's work on artificial life? Two major differences leap to the eye. First, Varela shifts the focus from the issue of the cell group's self-reproduction to that of its unity and boundary maintenance. As he (and Maturana) put it: "...for reproduction to take place there must be a unity to be reproduced: the establishment of the unity is logically and operationally antecedent to its reproduction."[36] In Langton's self-reproducing loops the boundary is a static construction, a shell within which the coding elements executing the cell's instructions for reproduction are encased. (In Varela's terminology, the shell is an allopoietic device.) In contrast, the formation of a self-enclosing boundary and its dynamic maintenance is what is most essential to Varela's model. A second difference is defined by the role of information-processing. In Langton's research the cellular automata are configured as information machines. Indeed, a specific configuration of information is what reproduces itself: both the "medium" and the "event" of this reproduction are constituted as information. For Varela, in contrast, the informational process (the calculation of output from input) is secondary, and necessarily subordinate to the dynamical processes that constitute and

[36] Varela, Maturana, and R. Uribe, "Autopoiesis: The Organization of Living Systems, Its Characterization and a Model," in BioSystems 5 (1974), 189. This article also contains the cellular automaton model.

maintain the cell group's unity (in a word, the processes that constitute its organization). Informational processes, which would include the cell group's reproduction, are allopoietic "sub-machines." As such, they do not define what it is that gives "life" to a living system. In his CA model this idea is illustrated by the formation and restoration of boundaries around a catalytic agent. It is the presence of the catalyst that initiates the dynamic process of self-organization, which does not involve information-processing except at the “lower level” of transition-state tables that determine whether particular cells are “on” or “off.”

What is to be made of these differences? In the proceedings of the “First European Conference on Artificical Life,” published under the title Toward a Practice of Autonomous Systems, Varela (with co-editor Paul Bourgine) spells out succinctly what he thinks Artificial Life research should be, and how his view differs from Langton's earlier definition of the field. The key concept for Varela is autonomy:

> Autonomy in this context refers to [the living's] basic and fundamental capacity to be, to assert their existence and to 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 suggests an enormous range of applications at all scales, from cells to societies. [37]

Whereas Langton seeks to "abstract the fundamental dynamical principles underlying biological phenomena, and recreat[e] these dynamics in other physical media —such as computers— making them accessible to new kinds of experimental manipulation and testing" (Langton et al., 1991, xiv), Varela thinks that "artificial life can be better defined as a research program concerned with autonomous systems, their characterization and specific modes of viability" (xi). Varela concedes that his view does not contradict

[37] Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificical Life, ed. Francisco J. Varela and Paul Bourgine (Cambridge, Mass.: The MIT Press, 1992), xi. In Chapter 6 I discuss several other examples and the influence of this orientation on contemporary robotics.

Langton's, but only makes it more precise. Yet, he also adds: "...it is by focusing on living autonomy that one can naturally go beyond the tempting route of characterizing living phenomena entirely by disembodied abstractions, since the autonomy of the living naturally brings with it the situated nature of its cognitive performances" (xi).

Although his tone is conciliatory, Varela unmistakably views Langton's work as "disembodied abstraction," and therefore of limited relevance, since it has no way of dealing with the "situated nature" of living systems and hence with their "cognitive performances." In fact, with the charge of “disembodied abstraction” Varela comes perilously close to associating Langton's methodology with that of symbolic computation, which had come to dominate Artificial Intelligence and cognitive science following the development of cybernetics and information theory in the 1950s. After a thirty-year period of dominance by a "research program emphasizing symbolic computations and abstract representations," Varela asserts, it is time to benefit from the re-discovery of connectionist models and neural networks. [38] Surprisingly, Varela seems to have completely forgotten (or willfully ignored) Langton’s explicit affiliations with these models and systems in nature that instantiate highly distributed, parallel processing. In any event, Varela concludes his introduction to <u>The Proceedings</u> by going one step further: having re-inscribed Artificial Life within the study of autonomous systems, he re-situates it within the evolving history of cognitive science. The latter, having started with classical cognitivism (based on symbolic computational) and having passed through connectionism, has now arrived at what Varela characterizes as "an <u>enactive</u> view of cognitive processes, which also places the autonomy of the system at its center and is thus naturally close to AL" (xvi). By "enaction," Varela means the "emergence" of a new world through the co-determination of a "structural coupling." However, by putting a

[38] See Chapter 5 for a detailed discussion of the shift from symbolic computation to neural net modeling.

strong emphasis on "emergence," a key term in Artificial Life research, Varela would seem to want to annex Artificial Life to a new or transformed version of his own work.[39]

Whatever the limits of the computational, information-processing approach, it must be said in response that it is not fair or even accurate to drive a wedge between this

approach and the study of emergence and self-organization, as Varela does by identifying the latter exclusively with successor paradigms in his history of cognitive science.[40] After all, the experimental production of emergence and self-organization has been an intrinsic aspect of research at the Santa Fe Institute (and the putatively "American" approach to Artificial Life) from its inception. What has always been essential to this research is not simply the computational approach, but rather, as Langton himself makes clear, the emergence of forms of non-symbolic computation in nonlinear processes and systems. Indeed, as we have seen, one of the most important aspects of Langton's work is that it resolves the apparent antimony between computation or information processing and a dynamical systems approach. (As we'll see in Chapter 4, this new form of "emergent computation" becomes the explicit focus of significant research in nonlinear systems theory). Suspiciously, however, Varela makes no mention of nonlinear systems. Instead, he continues to insist upon –indeed, to harp on—the sense in which living systems are the result of particular histories and contingencies. What is most wrong about the computational approach to life, he emphasizes at a subsequent European workshop on Artificial Life (held in San Sebastian, Spain, in 1993), is that it leaves out this contingency: "If it's silicon contingency, if it's tin can contingency, fine

---

[39] This is more clearly apparent in his book, Invitation aux sciences cognitives (Paris: Editions du Seuil, 1996, orig. pub. 1988), where Varela suggests that "enaction" supercedes "emergence."

[40] For details, see Invitation aux sciences cognitives and Chapter 5 below. In his brief "history" Varela conflates symbolic computation and information processing, which form the basis of the cognitivist paradigm. This stage is succeeded by connectionism and neural net theory, which, although instances of emergence and self-organization, are still committed to information processing and representation. Finally, with his own theory of "enaction," cognitive science arrives at a fully acceptable model.

with me. What you cannot abstract out, centrifuge out, is that kind of process or situation that only comes from history."[41]

One might retort that of course one never escapes from history, but at the same time, “life” never allows “history” to remain intact. What Varela seems to abhor above all else is --in Deleuze and Guattari’s terms-- that A-Life deterritorializes biological processes and reterritorializes them in a digital medium. But surely this can’t be totally unlike some of the “contingencies” that living systems have always confronted; indeed, the usefulness of D&G’s perspective –however limited their understanding of specific biological process—is that it suggests that life itself arose and gained a stronghold precisely through such decodings and recodings. In any case, it is not my purpose to adjudicate between Langton and Varela, only to provide a framework for further discussion of issues that their work necessarily foments. At the very least, the area of common ground should now be clear: neither body of work is governed or constrained by the opposition between machines and organic life, but gestures toward some yet to be conceptualized understanding of their difference and relationship. And while questions about life's origins, evolution, contingency and history are essential to both, these questions cannot be easily or directly addressed if we persist in following conventional understandings of computation and ignore the special features of nonlinear dynamical systems. This will become clearer in the next chapter, where these questions are reframed in relation to the shortcomings of Darwinian evolutionary theory on one side and the processes of emergence and self-organization in nonlinear dynamical systems on the other. This development in turn raises the question of whether the framework of "complex adaptive systems" can provide a new conceptual synthesis.

Lest these issues seem only theoretical, I want to conclude this chapter with a brief exploration of two convergent lines of research on immune systems. After

---

[41] Varela, quoted by Stefan Helmreich in Silicon Second Nature (Berkeley: University of California Press,

summarizing Varela's biological theory of the immune system, I shall consider a parallel development that arises from an interest in machine learning. This will lead to a brief discussion of computer models of the immune system, viruses and their relationship to Artificial Life. Again, one of the central questions raised by this convergence is whether the information-processing approach and the dynamical systems approach are actually or necessarily as opposed as Varela and others have claimed. The evidence suggests, contrarily, that for complex systems like the immune system (and we shall later examine others), information-processing and dynamic behavior can be very closely related, implicating and relaying each other in ways yet to be fully understood.

Silicon Immune Systems and Viral Life

For Varela, the immune system of a biological organism presents a clear and compelling example of an autopoietic system: it possesses an autonomous unity, operationally closed but structurally coupled to the outside by a triggering mechanism that communicates perturbations. What's more, it serves an essential cognitive function in that it is responsible for the organism's very identity at the molecular level. It is hardly surprising therefore that Varela has devoted much of his research to the immune system, presenting a first version of this view in Principles of Biological Autonomy (1979), and continuing to publish important articles well into the 1990s that update it. Beginning with this first theoretical formulation, Varela has always advocated an alternative to the classical understanding of the immune system as the body's primary defense against disease and infectious agents. According to the classical view, a wide variety of highly interactive cells, the lymphocytes (popularly known as "white blood cells"), possess and produce markers known as "antibodies" that either protrude from their surface or circulate freely. When such an antibody comes into contact with a foreign infectious agent or antigen the antibody bonds with it chemically, which leads to its destruction. If

1998), p. 225.

the infection is severe, the particular type of lymphocyte with the right chemical "key" or recognition device will immediately clone itself until it produces a veritable army of antibodies attacking the foreign intruder. Needless to add, this "military" model raises a number of questions. How do the body's lymphocytes learn how to recognize the millions of different antigens? Equally important, how does a lymphocyte "know" how to recognize its own body's cells, since in the classical view "recognition" is a chemical "locking-on" that destroys the other cell? Furthermore, what makes this mechanism fail in autoimmune disorders? Finally and most generally, how does the immune system maintain a "catalogue" of all known foreign intruders, past and present, and how does it produce new types of lymphocytes to counteract the new types of intruders?

In Varela's view, the classical theory attempts to answer these problems with a series of *ad hoc* proposals and hypotheses. A case in point is the "clonal selection theory," which proposes a Darwinian model to explain how the antibodies required for the body's maintenance have evolved over time. As Varela points out, this theory assumes that the body's antibody repertoire is initially incomplete, which doesn't fit with the known facts. Moreover, the theory postulates that those clones that would "recognize" (and destroy) self-molecules are missing; since there is no known genetic mechanism that can account for this, it is assumed that these clones are filtered out and destroyed at the embryonic stage. (One version of this theory, called "clonal deletion," postulates that such lymphocytes are removed in the thymus.) Thus, the recognition problem – how to distinguish between self and non-self at the level of molecular profiles--is not really resolved.[42]

Varela's solution to the problems posed by the classical model is to adopt a radically different perspective, doing away with both the military metaphor underlying

---

[42] See Francisco J. Varela and Mark R. Anspach, "The Body Thinks: The Immune System in the Process of Somatic Individualtion," in Materialities of Communication, ed. Hans Ulrich Gumbrecht and K. Ludwig Pfeiffer (Stanford: Stanford University Press, 1994), pp. 273-285.

the classical view and the information-processing model it assumes for its operation. (The former understands the maintenance of the body's identity at the molecular level as essentially a negative, wholly defensive reaction, while the latter views the body's immune response as a type of input-output relationship and therefore "externally determined.") Building on the work of Neils Jerne, who proposed that antibodies do not operate as separate, individual elements but as tightly meshed networks, Varela argues that the immune system is an autonomous network that must first be understood in positive terms. In his view the defensive reaction of the immune system is actually peripheral to its normal functioning. Its primary function, rather, is to maintain the body's molecular identity by regulating the levels of different cell types circulating throughout the entire system. Only when the invading antigens become so numerous as to perturb these regulatory functions does the immune system assume or fall back on its defensive posture. The problem, then, is not how the body identifies individual antigens but how it regulates levels of a whole range of interacting molecules. For this reason Varela prefers to think of the immune system as part of a larger, autonomous network that constitutes a complex ecology:

Like the living species of the biosphere, [lymphocytes] stimulate or inhibit each other's growth. Like the species in an ecosystem they generate an amazing diversity: the antibodies and other molecules produced by lymphocytes are by far (by a million fold) the most highly diversified molecular group in the body. They are therefore ideally qualified to ensure the constant change and diversity of other molecules in the body. (274)

This autonomous network functions as a dynamical system, with global emergent properties that enable it to track and remember the individual's molecular history:

In our view the IS [immune system] asserts a molecular self during ontogeny, and for the entire lifetime of the individual, it keeps a memory of what this molecular self is…. It is as a result of this assertive molecular identity that an individual who had measles in childhood is different from what he would have been had he not been in contact with the virus, or how an IS changes if the person switched from an omnivorous to a vegetarian diet. The IS keeps track of all this history, while defining and maintaining a sensorial-like interface at the molecular level. It must be stressed that the self is in no way a well-defined (neither pre-defined) repertoire, a list of authorized molecules, but rather a set of viable states, of mutually compatible groupings, of dynamical patterns.[43]

As a dynamical system, the immune network functions not by guarding and protecting boundaries between self and non-self, but by keeping different groups and sub-networks in states of dynamic equilibrium. In Varela's view, the immune system does not and cannot discriminate between self and non-self:

The normal function of the network can only be perturbed or modulated by incoming antigens, responding only to what is similar and to what is already present. Any antigen that perturbs the immune network is by definition an "antigen of the interior," and will therefore only modulate the ongoing dynamics of the network. Any element that is incapable of doing so is simply not recognized and may well trigger a "reflexive" immune response, that is, one produced by quasi-automatic processes that are only peripheral to the network itself. ("The Body Thinks," 283)

There are always antigens present in the network, just as there are always antibodies that attack other antibodies. The result is a ceaseless change in levels, quantities, and distributions, and thus of perturbations in an already existing network that is constantly (re)adjusting to itself. The most apt metaphor is not a military campaign intent upon vanquishing the enemy "other" but the dynamics of a weather pattern. The latter "never settles down to a steady–state, but rather constantly changes, with local flare ups and storms, and with periods of quiescence."[44] Yet the immune system is not exactly like a weather system either, since it "remembers" and draws upon its own history in order to function.

Indeed, the immune system's ability to remember and learn, that is, to evolve new pattern-recognition capacities, is precisely what made it of interest to J. Doyne Farmer and Norman Packard when they organized a conference devoted to "Evolution, Games,

[43] Francisco J. Varela, Antonio Coutinho, Bruno Dupire and Nelson N. Vaz, "Cognitive networks: Immune, Neural, and Otherwise," in Theoretical Immunology, Part Two (Redwood City CA: Addison-Wesley, 1988), p. 363.
[44] Alan Perelson makes this comparison in "Toward a Realistic Model of the Immune System," in Theoretical Immunology, Part Two (p. 396).

and Learning: Models for Adaptation in Machines and Nature."[45] One of the central questions they wanted to put on the table is how machines might learn to solve problems without having to be explicitly programmed to do so. Since biological systems –most evidently the immune system and brain—accomplish this task as part of their regular functioning, they offer privileged models for studying the underlying principles of biological computation. Since games are "highly simplified models of higher level human interaction," they too can provide access to "smart algorithms." But most important, as Farmer and Packard emphasize in their introduction to the conference proceedings, the behavior of these various adaptive systems is inevitably nonlinear and emergent:

> Adaptive behavior is an emergent property, which spontaneously arises through the interaction of simple components. Whether these components are neurons, amino acids, ants, or bit strings, adaptation can only occur if the collective behavior of the whole is qualitatively different from that of the sum of the individual parts. This is precisely the definition of the nonlinear. (vii)

Consequently, the approach to such nonlinear dynamical systems requires new syntheses, rather than further elaborations of the reductive approach, which breaks down complex processes into simpler component parts and processes. The most powerful and innovative tool available for the new synthetic approach is the computer, which provides a way of simulating adaptive systems too complex to model quantitatively. Although orders of magnitude simpler than "the brain or a complex organic molecule such as DNA" (viii), the digital computer's increasing speed, decreasing cost and wide availability have made it an essential part of what Farmer and Packard think of as an explosion of "new wave science," characterized by the synthetic approach and the crossing of conventional disciplinary boundaries.

In their own contribution to the conference, a paper entitled "The Immune System, Adaptation, and Machine Learning," Farmer, Packard and third co-author Alan

---

[45] The conference was held in 1984 at the Los Alamos National Laboratory. Proceedings were later published in Physica 22D (1986). Page numbers for citations will be inserted directly into the text.

S. Perelson proposed a dynamical model of the immune system simple enough to be simulated on a computer. Their central interest is the mechanism by which the immune system is capable of learning, memory and pattern-recognition, and specifically in the fact that "by employing genetic operators on a time scale fast enough to observe experimentally, the immune system is able to recognize novel shapes without preprogramming" (187). Because of its incredible combinatorial diversity, the immune system is able to generate very rapidly a large number of different types of antibodies (for a typical mammal like a human or a mouse, it is on the order of 10^7-10^8) capable of recognizing an even larger number of foreign molecules (estimates range as high as 10^{16}). A realistic model must be able to not only match this capacity but also provide a means by which the list of antibody and antigen types can constantly change, as new types are added and removed.

In the model proposed by Farmer, Packard and Perelson, both antibodies and antigens are represented by binary strings, which allow for either full or partial complementary matches. In this way, digital bit strings model "recognition" in natural immune systems, where matching is thought to occur when the molecular shape of the antibody's *paratope* "fits" with and thereby allows chemical binding to the antigen's *epitope*. Both paratopes and epitopes are sequences of amino acids, and thus complexly related, but researchers generally think of their relationship as that of a "lock and key." (The dynamics of the system are further complicated by the fact that antibodies also possess epitopes, which are recognized by other antibodies, and thus also participate in a self-regulating function.) Fortunately, in natural immune systems each antibody type (or "key") will fit a variety of antigen types (or "locks"). Yet this mechanism alone does not bridge the huge gap between the numbers of antibody and antigen types. In fact, since an organism never ceases to encounter both old and new antigens throughout its lifetime, its immune system must constantly "turn over," not only producing a large supply of antibody types effective against known and remembered antigens but also a repertory of new types that will “recognize” new antigens. In a human being, for example, the entire supply of lymphocytes dies and is replenished every few weeks. In order to remain effective, then, the system must produce a staggering combinatorial diversity of antibody types, and there are at least two mechanisms operating at different levels that insure this diversity. First, in the production of the cells in the bone marrow there is a constant reshuffling of the DNA that codes for antibody genes.[46] Second, when the lymphocytes themselves reproduce there is an exceptionally high mutation rate, much greater in fact than in any other body cell type.

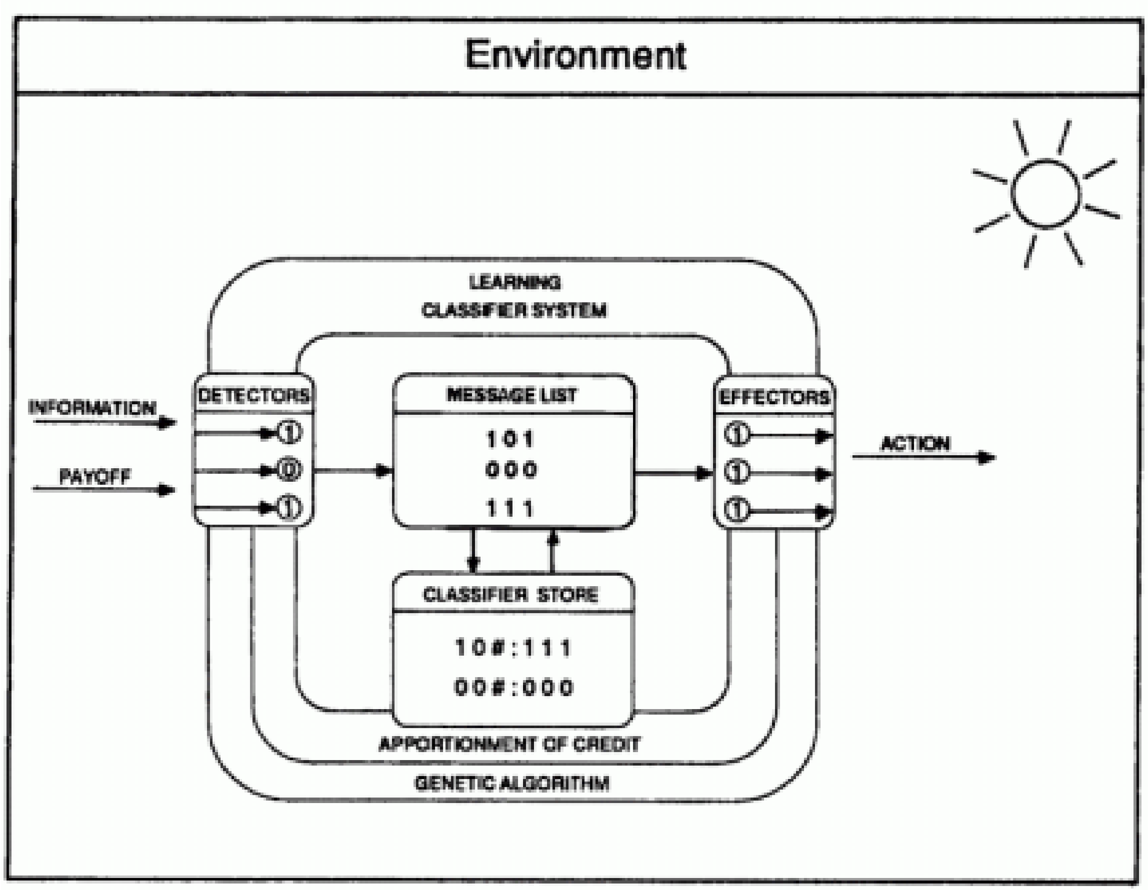
Of necessity, then, Farmer, Packard, and Perelson must make a number of simplifications; for example, they only attempt to model the actions of B-lymphocytes, and ignore the fact that natural immune systems also combine the actions of T-lymphocytes and macrophages. The central problem is how to mimic the production of this combinatorial diversity, clearly the immune system’s most essential and characteristic property. The genetic algorithm, developed by John Holland, provides the solution. By applying genetic operators like crossover, inversion and point mutation to both epitope and paratope bit-strings, a vast number of antibody types are generated that will basically comprise the model immune system. Antigens, on the other hand, can be generated either randomly or by design. These antigens are then repeatedly “presented” to the system, in both varying number and rate, to measure how well the system remembers (or how rapidly it forgets). Altogether, the total number of antibodies and antigens present at a given moment of time defines a single dynamical system, whose state will change as some of these antibodies and antigens interact and die, and both new and similar ones are added. The state of the system can thus be computed with a set of differential equations that allow new variables to be triggered into action as the

[46] See Farmer *et al*, p. 190, for the details. Basically, an antibody molecule consists of two polypeptide chains, each of which is coded for by a slightly different gene library. Thus there is a “hugh combinatorial amplification in the number of different antibody types that can be formed from a small number of gene libraries, each containing a limited number of genes.”

system evolves. In fact, the model consists of just such a set of equations. Having shown how they arrive at its formulation, the authors note the model's striking similarities to John Holland's classifier system, which they also "rewrite" as a set of differential equations in order to compare the two.

Both interesting in its own right and a highly influential new method of problem solving and learning in artificial intelligence, Holland's classifier system is certainly worth a brief summary. Basically, it is comprised of three components: a rule and message system, a credit system, and a genetic algorithm. In <u>Genetic Algorithms in Search, Optimization, and Machine Learning</u>, David E. Goldberg (one of Holland's former students) provides the following schematic diagram of a classifier system[47]:

[47] David E. Goldberg, <u>Genetic Algorithms in Search, Optimization</u>, and Machine Learning (Reading, Mass.: Addison-Wesley, 1989), p. 223.



A learning classifier system interacts with its environment.

As we can see from the diagram, information enters the system via detectors or sensors, where it is coded as a binary string and posted on a message list. It is then read by the classifiers, which are rules that take the form: “if <condition> then <action>”. If the <condition> is a match with the message, the <action> might be to post another message on the message list and/or to trigger an output action through an effector. The rules are also composed of binary strings, with an additional “wild card” marker (#) designating either 0 or 1. Thus, the message 0101 would match with rule 0##1. To illustrate, consider a simple system composed of the following four classifiers, with their outputs given after the colons (The example, which I have slightly modified, is actually Goldberg’s):

1. 01## : 0000

2. 00#0 : 1100
3. 11## : 1000
4. ##00 : 0001

Suppose the message 0111 from the environmental detector is posted on the message list. Since it matches with classifier 1, the latter would post the message 0000. This message in turn matches with classifiers 2 and 4, which would then post their messages, 1100 and 0001. Message 1100 now matches with classifiers 3 and 4, which posts messages 1000 and 0001. Of these two, only 1000 elicits a response: classifier 4 posts message 0001, which then elicits no response, terminating the process.

However, a classifier is not allowed simply to post a message when activated, as assumed above for purposes of exposition. In actuality it makes a bid to post a message, the efficacy of which is proportional to the classifier's strength as measured by the number of times its activation has led to an output action. Moreover, it must give credit to any and all classifiers whose activation has led to its own activation. In short, its own activation simply initiates a complex process of credit assignment, which then conditions its ability to participate further in a larger process. Holland has called this system of auction, payment, and reinforcement the "bucket-brigade algorithm," and the result is what Goldberg likens to "an information economy where the right to trade information is bought and sold by classifiers…[which] form a chain of middlemen from information manufacturer (the environment) to information consumer (the effectors)" (225).

With such rigorous and competitive demands on the rules or classifiers, one may well wonder how the system can be supplied with rules that work at all, much less with high efficiency. This is where the third part of the classifier system, the genetic algorithm, comes in, for it provides the means to constantly generate new and better rules that can be injected into the system. This third part works in conjunction with the credit assignment sub-system, which separates the good rules (i.e., those that perform effectively) from those that do not. By applying the processes of crossover and mutation to these good rules, new and better ones are bred, which are then added to the population of rules to have their performances evaluated and winnowed in turn, as the whole system gradually turns over. Needless to add, as an alternative to the serial processing of traditional expert systems, the parallel rule activation of the classifier system avoids the bottlenecks of the latter and allows multiple activities to be coordinated simultaneously. While this makes for a much faster hardware implementation, the real advantage is that it results in a machine that can actually learn and adapt to changing information.

Although the classifier system is clearly an information-processing device, the fact that the rules operating within it are changing over time means that it functions as a nonlinear dynamical system. As Farmer, Packard and Perelson demonstrate, both the classifier and model immune systems are strongly nonlinear, and the equations for computing their changing behavior take the same basic form:

Δx = internal interactions + driving – damping

Although the precise form these terms take depends upon how the interactions as well as the driving and damping forces all influence one another, the general form of the equation is often seen in biological phenomena – in particular, the authors mention coupled autocatalytic reactions and the Lotke-Volterra equations for population dynamics. Yet these similarities are hardly surprising, given that both their model of the immune system and Holland's classifer system mimic parallel computation in natural (i.e., biological) systems. Indeed, after an extended comparison of the two systems, Farmer, Packard and Perelson conclude by affirming the superiority of such parallel computational systems over standard serial Turing machines. Not only does their model provide insight into the internal operations of real immune systems but the correspondences between their model and the classifier system reinforces the growing sense "that generalized versions of our model may be capable of performing artificial intelligence tasks" (203). Indeed, parallel processing and the use of genetic algorithms in classifer systems have since become part of the repertory of contemporary AI.

In several of his later essays Varela acknowledges Farmer, Packard and Perelson's work on the immune system as a dynamical system and its affinities with Holland's classifier system. At the same time, however, Varela fails to take into account the implications of this synthesis of an information-processing and dynamical systems approach. Since I will pursue some of the implications of this line of research in the next chapter, I would like to bring this chapter to a close with a brief look at some related research on computer immune systems.

That there is research on computer immune systems – which stands in clear contrast to the commercial development of anti-virus software--is of course a highly significant fact. Needless to say, the development of both presupposes an environment of proliferating and increasingly sophisticated computer viruses. These viruses began to appear in the early 1980s, and it was only a matter of time before they would be considered in relation to –and even as forms of--Artificial Life. Although no explicit mention of computer viruses appears in the proceedings of the first Artificial Life Conference in 1987, Langton's bibliography contains a reference to A. K. Dewdney's Scientific American article, "Computer Recreations: A Core War Bestiary of Viruses, Worms and Other Threats to Computer Memories," published in 1985.[48] Dewdney had earlier invented a computer game called "Core Wars" in which two sets of digital organisms within a virtual world attack each other's machine instructions, and Langton invited him to the conference to judge an A-life "contest" mainly intended to amuse and entertain the participants. In his book Artificial Life, Stephen Levy rightly devotes a number of pages to Dewdney, as well as to Fred Cohen, who while a computer science

[48] In Artificial Life II (Addison-Wesley, 1992), however, Langton will publish Eugene Spafford's article "Computer Viruses – A Form of Artificial Life?" While Spafford acknowledges that science has much to

graduate student at USC in 1983 had written a virus of two hundred lines of code that could invisibly give him system administrator privileges on a Unix machine. Cohen was one of the first "professional" experimenters with computer viruses –in 1987 he would publish the results of his experiments in the reputable journal Computers and Security -- but the line of demarcation was not always clear. When the twenty-one year old Cornell student Robert Morris released his self-replicating "Internet Worm" in the same year, quickly paralyzing some six thousand computers, his actions not only created panic and hysteria but eventually resulted in the establishment of a new panoply of legal measures and law enforcement agencies. Cohen's own laconic response was to remark that Morris had just set the world's record for high-speed computation. Indeed, in those inchoate times both controlled and uncontrolled "experiments" with forms of artificial life *avant la lettre* could only be worrying and problematic to representatives of official scientific culture. Levy quite correctly sets the two perspectives side by side:

> **During the period that young Morris and other unauthorized experimenters were blithely releasing predatory creatures in the wild [i.e., onto early versions of the Internet], Cohen and other serious researchers were consistently being refused not only funding but even permission to conduct experiments in computer viruses. As a result, the creations of willful criminals and reckless hackers were for years the most active, and in some ways the most advanced, forms of artificial life thus far. (324)**

Whether intentionally or not, Levy leaves us with the impression that a whole new realm of artificial life is burgeoning around us, some of which is scientifically "authorized" and officially sanctioned, while other forms constitute an unauthorized but no less fertile "underside." While no doubt this represents the "official" position, the boundary line it assumes is perceptibly misleading and ever-shifting. For example, "outsider" Mark A. Ludwig's book, Computer Viruses, Artificial Life and Evolution (1993), contains astute technical and philosophical discussions of Artificial Life, while also providing computer code for experimenting with a variety of real viruses. Numerous practices within the legally sanctioned sector of this realm, moreover, suggest how

---

learn from studying computer viruses, he is disturbed that their "origin is one of unethical practice" (744).

problematic the term “authorized” actually is. Within a few years of its appearance, Dewdney’s computer game “Core Wars” was taken up and rewritten by a small group of scientists led by Steen Rasmussen, in order to investigate “the emergence and evolution of cooperative structures in computational chemistry.”[49] And later the computer game “SimLife” (produced by Maxis in 1993) was widely hailed as an important scientific prototype for studying the evolution of digital organisms. More recently, software inspired by A-life research has been used in turn to create sophisticated computer games like “Evolva” and “Creatures.” Meanwhile, in a striking example of advanced A-life research, John R. Koza has experimented with computer programs that spontaneously emerge, self-replicate and evolve self-improving versions.[50] However, what “authorizes” these activities is neither the creative individual nor the permissive institution but the performative capacity of machinic, an-organic life, as it is explored, “worked” and further extended along lines of continuous variation by silicon probeheads who resemble in many ways the nomadic metallurgists of an earlier era.

In any case, pioneer researchers in the field of computer immune systems must necessarily view the world of computers as a site where new forms of viral life are now emerging. In an article published in Artificial Life IV, Jeffrey O. Kephart argues that current anti-virus techniques are doomed to fail, and must eventually be replaced by a biologically-inspired immune system for computers.[51] Although an important step, Kephart’s model still assumes the military understanding of the immune system that Varela contests. In contrast, a potentially more fruitful approach is taken by Stephanie Forrest, who conceives of the world of computers as having many of the properties of a living ecosystem, populated with “computers, with people, software, data, and

---

[49] See S. Rasmussen, C. Knudsen, R. Feldberg and M. Hindsholm, “The Coreworld: emergence and evolution of cooperative structures in a computational chemistry,” in Emergent Computation, ed. Stephanie Forrest (Cambridge, Mass. The MIT Press, 1991), pp. 111-134.
[50] See John R. Koza, “Artificial Life: Spontaneous Emergence of Self-Replicating and Evolutionary Self-Improving Computer Programs,” in Artificial Life III, ed. Christopher G. Langton (Reading, Mass.: Addison-Wesley, 1994), pp. 225-262.

programs…." [52] In "Principles of a Computer Immune System,"[53] Forrest lists some twelve organizing principles of a biological immune system – many of which, like autonomy, adaptability, and dynamically changing coverage, while not going as far as Varela, move beyond the strictly defined military model. Forrest argues that these principles must be incorporated as design principles if a computer immune system is to function. However, if the objective is "to design systems based on direct mappings between system components and current computer system architectures" (79), then the latter will have to be radically modified. One possible architecture, she suggests, would be something like an equivalent "lymphocyte process," comprised of lots of little programs that would query other programs and system functions to determine whether they were behaving normally or not. But they would also have to monitor each other, "ameliorating the dangers of rogue self-replicating mobile lymphocytes" (80). Just how feasible this approach will turn out to be remains difficult to say, and Forrest is rightly cautious. In fact, she is acutely aware of the limitations of "imitating biology," since biological organism and human-made computers have very different methods and objectives.

But here we may be confronting the limits of a conceptual contradiction between the computer as a tool or medium which we can control and the computer as part of an ecosystem in which it cannot function unless it is given more life-like capacities that will put it outside our control. Perhaps human computational and communicational activities will eventually have to be made more like biological exchanges if a fully functional computer immune system is to be constructed –or rather, and more likely-- evolved. In any case the theoretical first steps towards the evolution of such systems is to be found in the study of what are now known as complex adaptive systems. One of the most interesting features of such systems is that they emerge in both nature and culture. It is

---

[51] Jeffrey O. Kephart, "A Biologically Inspired Immune system for Computers," in Artificial Life IV, ed. Rodney A. Brooks and Pattie Maes (Cambridge, Mass.: MIT Press, 1994), pp. 130-139.
[52] Quoted by Lesley S. King, "Stephanie Forrest: Bushwacking Through the Computer Ecosystem," SFI Bulletin Vol. 15, No. 1(Spring 2000).
[53] Presented at New Security Paradigms Workshop, Langdale, Cumbria UK, 1998.

also no accident that they should come into view –that is, be identified as such– about the same time that Artificial Life appears. This convergence will be the concern of the next chapter.