IS ARTIFICIAL LIFE A CONTRADICTION IN TERMS? © Dan Bruiger 2012

Is an organism a machine? Under what circumstances can a machine *become* an organism? Clear answers to these questions are important to evaluate the feasibility and risks of artificial life, artificial intelligence, and nanotechnology. I will argue that an organism is not a machine because it is not an artifact. The answer to the second question hinges on an understanding of how analytical thought and the philosophy of mechanism can mislead us into thinking that natural reality can be formally exhausted in thought and recreated as artifact. A machine can become an organism only by creating itself from the bottom up, not by being created from the top down, through formal analysis or reverse engineering. Moreover, it cannot be both autonomous *and* fully subject to human control. This fact presents us with a watershed choice, based on a fundamental tradeoff: create intelligent artifacts as tools of human intent; or foster autonomous artificial organisms, which may elude human control and pose a threat to all life.

Much of the current optimism of genetic engineering rests on the flawed assumption that organisms *are* machines, whose genetic program is their blueprint. No natural thing is literally a machine, however, because natural reality is *found*, not *made*. The quest to engineer the artificial organism rests on the theoretical possibility to exhaustively analyze the natural one. However, no natural thing *can* be exhaustively analyzed. Only things that were first encoded (messages) can be decoded. Only things that were designed can be reverse-engineered. The artifact produced in analysis is not the natural thing it models, but an idealized formalism.

The implicit idealist faith behind theoretical modeling, and the ideal of perfect simulation, is that each and every property of a thing can be represented with precision. A 'property', however, is a human artifact, an assertion that disregards a potential infinity of other assertions. The collected properties of a natural thing do not constitute it, although they do constitute an artifact. Imitating natural organisms is an effective way to produce artifacts, not true artificial organisms.

Unlike an organism, a machine is an artifact by definition. And, unlike a machine, an organism is not a product of definition. While a machine might be inspired by natural systems, someone designed it. It has a precisely "true" set of well-defined parts, intended by its fabricators. It can be dismantled into this same set of parts by reversing the process of construction: the process of design is the reverse of analysis. The mechanistic view of the cosmos assumes that the universe itself is a machine that can be analyzed into its true parts, in the way that an engine can be taken apart and reassembled. However, one is always only guessing at the parts of any natural system or thing. The basic problem for those who want to engineer or reverse-engineer "life" is that we simply did not make the organism in the first place. The basic problem for those who want to develop artificial "intelligence" is that natural intelligence is an embodied phenomenon. This means that it follows its own priorities, whereas artificial intelligence follows priorities defined by human engineers.

While a machine is a closed reversible system, an organism is not. One cannot truly understand the functioning and behavior of even the simplest creature, or its genetic basis, without grasping the complex interactions with environments that are their source and to which they refer. Just as a computer program draws not only upon logic and the

mechanics of the computer, but also upon the semantically rich environment of the programmer, so the developing embryo, for example, does not simply unfold according to a program spelled out in genes, but through complex chemical interactions with the uterine environment and beyond. The genetic "program", in other words, is not a self-contained syntactic system, but is rich in references that extend indefinitely outside itself. The organism is both causally and—if I may say so—*intentionally* connected to the rest of the universe. Simply to identify genetic units of information is not to understand the genetic "code", any more than recognizing units of a foreign language as words implies understanding their meaning.

Simulation presumes that natural processes, objects, and actions can be taken apart in thought, abstracted in a formalism, then reconstructed in an artificial version from the formalism serving as a design. The flight of airplanes resembles the flight of birds and insects only metaphorically, however. A model airplane that flies may perfectly simulate a real airplane because, though reduced in size, it *is* a real airplane. In contrast, an artificial insect or bird is not a real one, even should it fly. Turing's Universal Machine (the digital computer) can simulate any other *machine* exhaustively. But whether a machine, program, artifact, or theoretical model can exhaustively simulate an organism, or *any* aspect of natural reality, is quite another question.

What makes perfect simulation naively seem feasible may be the characteristic "chunking" involved in language and thought, whereby a rose is a rose is a rose. But there are many varieties of rose and every individual blossom is unique. A baseball player and a pitching machine are both called pitchers; language permits us to say they perform the "same" action. But the device only crudely imitates the human, no matter how accurately it hurls the ball. "Pieces" of behavior and natural entities can only conventionally be identified. Common sense recalls the differences between real and artificial flowers, or between the intricate human movements of throwing and the mechanical hurling of the ball. Yet, the concept of simulation rests on obscuring such distinctions, by conflating all that can pass semantically under a given rubric.

One goal of Artificial Intelligence, of course, is to defy common sense by creating ever more sophisticated robotics to mimic aspects of human behavior. There may soon be robots that can achieve higher baseball scores than human players, just as there are computers that can beat humans at chess and "Jeopardy". Machines may be made to simulate human (or any) behavior *as specifically defined in some formalism*. Yet, the machine is enacting the formalism, not duplicating the human being. It is fundamentally misguided to argue for the identity of robot and human on the basis of their common behavior, no matter how fine the resemblance, for the commonality can never be exhaustively established. All that can be known with certainty is that the machine follows its program. To assume that the human organism follows an identical program is simply circular reasoning.

Similarly, it is a mistake to think that the essence of any natural thing can be exhaustively modeled, or captured in a program or blueprint for its construction. The formalism, model, or program is the semantic bottleneck through which the whole being or behavior of a natural thing must pass in order to be simulated. One thing is said to simulate another when they both embody a common formalism. This can work perfectly well for two artifacts such as the airplane and its toy version, because they are alternative

realizations of the same design, two tokens of a common ideal. However, it cannot be known, through some finite procedure of checking, that the being of a natural object is exhausted in some formalism abstracted from it. It is a mistake to take this formalism for its blueprint or essence, in the sense that the aeronautical engineer's design is the blueprint defining the essence of the flying machine. The natural object is a *found object*, not an invention constructed from design. The blueprint is constructed after the fact, inferred from an analysis that can never be guaranteed complete. The mechanist fallacy is to believe that it is possible to replicate a natural object by first formalizing its structure and behavior and then constructing an artifact from that as a design. The artifact *will* instantiate the design, of course. But it will *not* duplicate the natural object, any more than an airplane duplicates a bird.

While an organism is not a machine, yet it may be possible for a machine to be an organism. Paradoxically, however, the key is that it cannot be an *artifact*. What begins as a product of design must bootstrap itself into the self-design that characterizes organism. An organism is self-defining, self-assembling, self-maintaining, self-reproducing. In a word, it is *autopoietic*.[i] Under certain conditions, an artifact might self-assemble, self-maintain, even self-replicate—as in the case of von Neumann's theoretical self-replicators. As things stand, however, it would not be self-defining. In order to become an organism, it must acquire its own purposes, so to speak. The way this has come about in Nature is through natural selection over many generations. The priorities or purposes of the organism are established through a contest for survival. Events in the world *matter* to the organism, whose welfare depends on them. While a machine exhibits only the intentionality of its designers, the organism derives its own intentionality from participation in evolutionary contests, through a long history of interactions with environments consisting significantly of other co-agents. Paradoxically, the organism has life by virtue of its mortality and has individuality by virtue of its kind's genetic success.

Organisms develop within rich and highly specific environmental contexts. The natural organism stands on the shoulders of its ancestors, moreover, while the self-replicating automaton may be stranded on the shoulders of its programmers, with their short, impatient life spans. It should not be assumed that technological development is equivalent to biological evolution, nor based on similar premises. Technological development as we know it is an extension of human purposes, employing top-down design from an anthropocentric point of view. It may appear to mimic biological evolution insofar as there is a process of selection within the marketplace of ideas. Yet, while a designed system (a machine) may be physical, it is not *embodied*. Embodiment is the outcome of self-definition, acquired through adaptation in an evolutionary contest. It implies a network of relationships and their history. Unlike the machine, the organism has its own priorities and its own point of view, which do not coincide with those of human designers. No machine, no matter how complex, as yet has its *own* purposes, not endowed by design. For, this has never been explicitly the goal of its human developers, and never should be. The tinkering of mad scientists aside, the serious application (and funding) of robotics requires machines that serve human goals and remain under control, even when fast and cheap.

Artificial life, with its own intentionality, *might* effectively emerge through evolutionary games, much as natural life did—games such as are currently explored in

the field of computer simulation known as A-Life. At some point, such computational entities might be coupled with physical entities, such as self-replicating nanomachines, giving them a physical basis to develop embodiment in real-world contests through reproduction. This, however, would be an extension of natural selection, which would then include artificial creatures competing with natural ones. Such new creatures would have real autonomy and would no longer be under direct human control; we would not be able to define their priorities. At best, we could exercise the kind of control that we presently assert over natural organisms, and with similar uneven results. If artificial organisms prove superior in the real world, then efforts to control them would be all the more difficult, or futile.

In the end, we must ask what is the *point* of facilitating the evolution of true artificial life, aside from the sheer claim to have done it? Even if superintelligent machines could manage further intellectual and technological progress better than we bumbling humans, they might do *everything* so well that there would be no further need or place for human life, except perhaps in a zoo. The creation of superintelligence might be the last human mistake.

If we wish to retain control, and human hegemony on the planet, there will be necessary limits to the autonomy of technology, which means limits on the scope of its intelligence—if by this we mean the sort of general and self-interested intelligence we expect from living creatures. This is a political as well as a design choice. (There are indeed people—transhumanists or posthumanists—who embrace the prospect of a superior artificial successor to the human race.) It is fortunate that the conventional top-down design approach can *only* lead to useful, if dangerous tools. Yet, even these could reach a state of sophistication beyond human control, whether or not they constitute autonomous organisms. We may simply create a technological environment that is too complex for our limited intelligence to manage, the so-called "singularity".

I do not deny that machines of the future *could* be autonomous and even sentient; but if they are, I deny that they will be *machines* in the presently understood sense—that is, tools of human purpose, disembodied intelligence with no intentionality of their own. There is little doubt that new entities theoretically could self-create under the right circumstances—after all, life did it at least once. But there is every reason to doubt that, if successful in competing in the biosphere, these would remain within the sort of human control exerted over either conventional machines or natural organisms. We must distinguish clearly between machines that are extensions of their designers' motivations and autonomous machines that are creatures with their own motivations and survival instincts. The latter would be no more controllable than an invading race of space aliens. We must choose—while we still can—between a technology that serves us and remains under our thumb and the dubious accomplishment of siring new forms of being that could destroy the biosphere, supplant us, or bring us into servitude to them.

[i] A term coined by biologists Humberto Maturana and Franciso Varela in *Autopoiesis and Cognition* 1973