Edge

Summer Reading from the Archive

THE EMERGENT SELF

A Conversation with Francisco Varela

[June 2001]

> *Why do emergent selves, virtual identities, pop up all over the place, creating worlds, whether at the mind/body level, the cellular level, or the transorganism level? This phenomenon is something so productive that it doesn't cease creating entirely new realms: life, mind, and societies. Yet these emergent selves are based on processes so shifty, so ungrounded, that we have an apparent paradox between the solidity of what appears to show up and its groundlessness. That, to me, is the key and eternal question.*

Introduction

Francisco Varela, an experimental and theoretical biologist, studied what he termed "emergent selves" or "virtual identities." His was an immanent view of reality, based on metaphors derived from self-organization and Buddhist-inspired epistemology rather than on those derived from engineering and information science. He presented a challenge to the traditional AI view that the world exists

independently of the organism, whose task is to make an accurate model of that world—to "consult" before acting. His non-representationalist world—or perhaps "world-as-experienced"—has no independent existence but is itself a product of interactions between organisms and environment. He first became known for his theory of autopoiesis ("self-production"), which is concerned with the active self-maintenance of living systems whose identities remain constant while their components continually change.

Varela is tough to categorize. He was a neuroscientist who became an immunologist. He was well informed about cognitive science and was a radical critic of it, because he was a believer in "emergence"—not the vitalist idea promulgated in the 1920s (that of a magical property that emerges inexplicably from lower mechanical operations) but the idea that the whole appears as a result of the dynamics of its component parts. He thought that classic computationalist cognitive science is too simplemindedly mechanistic. He was knowledgeable and romantic at the same time. I sat down to talk to him in 1984 (in both Connecticut and Paris) to interview him for my book, *The Third Culture*.

As theoretical biologist Stuart Kauffman noted at the time, "Francisco Varela is amazingly inventive, freewheeling, and creative. There's a lot of depth in what he and Humberto Maturana have said. Conversely, from the point of view of a tied-down molecular biologist, this is all airy-fairy, flaky stuff. Thus, there's the mixed response. That part of me that's tough-minded and critical is questioning, but the other part of me has cottoned on to the recent stuff he's doing on self-representation in immune networks. I love it."

—JB

—

FRANCISCO VARELA (1946–2001) was a Chilean biologist, philosopher, cybernetician, and neuroscientist who, together with his mentor Humberto Maturana, is best known for introducing the concept of autopoesis to biology, and promoting dialogue between science and Buddhism. He was director of research at the Centre National de Recherche Scientifique, and professor of cognitive science and epistemology at the École Polytechnique, in Paris. He is the author of *Principles of Biological Autonomy*, co-author with Humberto D. Maturana of *Autopoiesis and Cognition: The Realization of the Living* and *The Tree of Knowledge*, and with Evan Thompson and Eleanor Rosch of *The Embodied Mind*.

THE EMERGENT SELF

I guess I've had only one question all my life. Why do emergent selves, virtual identities, pop up all over the place creating worlds, whether at the mind/body level, the cellular level, or the transorganism level? This phenomenon is something so productive that it doesn't cease creating entirely new realms: life, mind, and societies. Yet these emergent selves are based on processes so shifty, so ungrounded, that we have an apparent paradox between the solidity of what appears to show up and its groundlessness. That, to me, is a key and eternal question.

As a consequence, I'm interested in the nervous system, cognitive science, and immunology, because they concern the processes that can answer the question of what biological identity is. How can you have some kind of identity that simultaneously allows you to know something, allows cells to configure their own relevant world, the immune system to generate the identity of our body in its own way, and the brain to be the basis for a mind, a cognitive identity? All these mechanisms share a common theme.

I'm perhaps best known for three different kinds of work, which seem disparate to many people but to me run as a unified theme. These are my contributions in conceiving the notion of autopoiesis—self-production—for cellular organization, the enactive view of the nervous system and cognition, and a revising of current ideas about the immune system.

Regarding the subject of biological identity, the main point is that there is an explicit transition from local interactions to the emergence of the "global" property—that is, the virtual self of the cellular whole, in the case of autopoiesis. It's clear that molecules interact in very specific ways, giving rise to a unity that is the initiation of the self. There is also the transition from nonlife to life. The nervous system operates in a similar way. Neurons have specific interactions through a loop of sensory surfaces and motor surfaces. This dynamic network is the defining state of a cognitive perception domain. I claim that one could apply the same epistemology to thinking about cognitive phenomena and about the immune system and the body: an underlying circular process gives rise to an emergent coherence, and this emergent coherence is what constitutes the self at that level. In my epistemology, the virtual self is evident because it provides a surface for interaction, but it's not evident if you try to locate it. It's completely delocalized.

Organisms have to be understood as a mesh of virtual selves. I don't have one identity, I have a bricolage of various identities. I have a cellular identity, I have an immune identity, I have a cognitive identity, I have various identities that manifest in different modes of interaction. These are my various selves. I'm interested in gaining further insight into how to clarify this notion of transition from the local to the global, and how these various selves come together and apart in the evolutionary dance. In this sense, what I've studied, say, in color vision for the nervous system or in immune self-regulation are what Dan Dennett would call "intuition pumps," to explore the general pattern of the transition from local rules to emergent properties in life. We have at our disposal beautiful examples to play around with, both in terms of empirical results and in terms of mathematics and computer simulations. The immune system is one beautiful, very specific case. But it's not the entire picture.

My autopoiesis work was my first step into these domains: defining what is the minimal living organization, and conceiving of cellular-automata models for it. I did this in the early 1970s, way before the artificial-life wave hit the beach. This work was picked up by Lynn Margulis, in her research and writings on the origins of life, the evolution of cellular life, and, with James Lovelock, the Gaia hypothesis. Humberto Maturana and I invented the idea of autopoiesis in 1970. We worked together in Santiago, during the Socialist years. The idea was the result of suspecting that biological cognition in general was not to be understood as a representation of the world out there but rather as an ongoing bringing-forth of a world, through the very process of living itself.

Autopoiesis attempts to define the uniqueness of the emergence that produces life in its fundamental cellular form. It's specific to the cellular level. There's a circular or network process that engenders a paradox: a self-organizing network of biochemical reactions produces molecules, which do something specific and unique: they create a boundary, a membrane, which constrains the network that has produced the constituents of the membrane. This is a logical bootstrap, a loop: a network produces entities that create a boundary, which constrains the network that produced the boundary. This bootstrap is precisely what's unique about cells. A self-distinguishing entity exists when the bootstrap is completed. This entity has produced its own boundary. It doesn't require an external agent to notice it, or to say, "I'm here." It is, by itself, a self-distinction. It bootstraps itself out of a soup of chemistry and physics.

The idea arose, also at that time, that the local rules of autopoiesis might be simulated with cellular automata. At that time, few people had ever heard of cellular automata, an esoteric idea I picked up from John von Neumann—one that would be made popular by the artificial-life people. Cellular automata are simple

units that receive inputs from immediate neighbors and communicate their internal state to the same immediate neighbors.

In order to deal with the circular nature of the autopoiesis idea, I developed some bits of mathematics of self-reference, in an attempt to make sense out of the bootstrap—the entity that produces its own boundary. The mathematics of self-reference involves creating formalisms to reflect the strange situation in which something produces A, which produces B, which produces A. That was 1974. Today, many colleagues call such ideas part of complexity theory.

The more recent wave of work in complexity illuminates my bootstrap idea, in that it's a nice way of talking about this funny, screwy logic where the snake bites its own tail and you can't discern a beginning. Forget the idea of a black box with inputs and outputs. Think in terms of loops. My early work on self-reference and autopoiesis followed from ideas developed by cyberneticists such as Warren McCulloch and Norbert Wiener, who were the first scientists to think in those terms. But early cybernetics is essentially concerned with feedback circuits, and the early cyberneticists fell short of recognizing the importance of circularity in the constitution of an identity. Their loops are still inside an input/output box. In several contemporary complex systems, the inputs and outputs are completely dependent on interactions within the system, and their richness comes from their internal connectedness. Give up the boxes, and work with the entire loopiness of the thing. For instance, it's impossible to build a nervous system that has very clear inputs and outputs.

The next area of significant work involves applying the logic of the emergent properties of circular structures to look at the nervous system. The consequence is a radical change in the received view of the brain. The nervous system is not an information-processing system, because, by definition, information-processing systems need clear inputs. The nervous system has internal, or operational, closure. The key question is how, on the basis of its ongoing internal dynamics, the brain configures or constitutes relevance from otherwise nonmeaningful interactions. You can see why I'm not really interested in the classical artificial-intelligence and information-processing metaphors of brain studies. The brain can't be understood as a computer, in any interesting sense, and I part company with the people who think that the brain does rely on symbolic representation.

The same intuitions cut across other biological fields. Deconstruct the notion that the brain is processing information and making a representation of the world. Deconstruct the militaristic notion that the immune system is about defense and looking out for invaders. Deconstruct the notion that evolution is about optimizing fitness to live in the conditions present in some kind of niche. I haven't been

directly active in this last line of research, but it's of great importance for my argument. Deconstructing adaptation means deconstructing neo-Darwinism. Steve Gould, Stuart Kauffman, and Dick Lewontin, each in his own way, have spelled out this new evolutionary view. Lewontin, in particular, has much appreciated the fact that my work on the nervous system mirrors his work with evolution.

My fourth area of concentration—the most recent one—consists of using the same concepts to revise our understanding of the immune system. Just as conventional biology understood the nervous system as an information-processing system, classic immunology understands immunology in military terms—as a defense system against invaders.

I've been developing a different view of immunology—namely, that the immune system has its own closure, its own network quality. The emergent identity of this system is the identity of your body, which is not a defensive identity. This is a positive statement, not a negative one, and it changes everything in immunology. In presenting immunology in these terms, I'm creating a conceptual scaffolding. We have to go beyond an information-processing model, in which incoming information is acted upon by the system. The immune system is not spatially fixed, it's best understood as an emergent network.

I've also carried out empirical work corresponding to these intuitions. These ideas are incarnated into new experiments, and provide new results. For example, in classical immunology you were dealing with an external response system that was always watching out for invaders. If this made sense, the system would shrink to nothing if there were no invaders. Yet when mice are raised in milieus free from external challenge, their immune systems are normal!

Classical medicine remains baffled by the spectrum of diseases known as autoimmune diseases. Why? Because autoimmune disease is outside the paradigm of immunology. There's nothing to vaccinate against; there's no bacteria coming from outside. It's something that the system does to itself. AIDS is a dramatic case of the deregulation of this coherent emergent property, much like ecological dysfunctioning. People think AIDS is an infection. This is, of course, true, but not true in the sense that once the system is infected with AIDS it triggers a condition of self-destruction of the immune system. HIV triggers a deregulation, which then amplifies itself and becomes its own nightmare. Thus, when you look at the urine of an AIDS-infected patient, less than 5 percent of the dead lymphocytes are HIV-infected.

This is typical of an autoimmune condition: the system eats itself up. Consequently, it's beginning to dawn on people that looking for AIDS vaccines is a complete waste of time. From my point of view, the right approach is first to understand the nature of this global regulation. One hint of how to do this is to look for ways to reconnect the system. In this regard, autoimmune diseases are seen as a deregulation, a condition that cries for more connectedness, rather than as a condition susceptible to treatment with a vaccine. For example, look at drug addiction in terms of a social disease: Drug addicts are in some sense an autoimmune disease of society, because they end up destroying segments of society. What those people need is to be given support, jobs, and family care; you reconnect them back into the society. One approach we study is to provide new, normal antibodies that help to re-create the network. We are researching more sophisticated ways of doing this, but we need to have a pointer on where to go. Vaccines are not the answer.

I'm interested in establishing empirical correlations between a long-standing interest in Buddhist practice and scientific work. Western tradition has avoided the idea of a selfless self, of a virtual self. This egolessness, or selflessness, is truly the core of Buddhism. Over the past two thousand years, the Buddhists have developed philosophical, phenomenological, and epistemological sophistication, and they have invoked this intuition in a very hands-on way. We can use these insights much like people in the Renaissance used Greek philosophy to try to understand the science of Galileo.

Buddhism is a practice, not a belief, and every Buddhist is, in some way, lay clergy—involved in the way a scientist is involved in his or her work, or in the way a writer's mind is involved in writing, present in the background, all the time. People today have the leisure and sophistication to practice what before was only practical for monks. Buddhism affects Western culture through the individuals who practice it, through people who occasionally take it up as an escape. Buddhist ideas are prevalent throughout our culture—in physics and biology, for example, the basic ideas are Buddhism in disguise.

My view of the mind has been influenced by my interest in Buddhist thought. Buddhists are specialists in understanding this notion of a virtual self, or a selfless self, from the inside, as lived experience. This is what fascinates me about that tradition. Dan Dennett, incidentally, has come to the same conclusion in his own way. But while Dan focuses on the cognitive level, my own approach is to think about several biological levels, as I have mentioned, perhaps because I'm influenced by the broad idea of non-representationalist knowledge. In my reality, knowledge coevolves with the knower and not as an outside, objective representation.

I see the mind as an emergent property, and the very important and interesting consequence of this emergent property is our own sense of self. My sense of self exists because it gives me an interface with the world. I'm "me" for interactions, but my "I" doesn't substantially exist, in the sense that it can't be localized anywhere. This view, of course, resonates with the notions of the other biological selves I mentioned, but there are subtle and important differences. An emergent property, which is produced by an underlying network, is a coherent condition that allows the system in which it exists to interface at that level—that is, with other selves or identities of the same kind. You can never say, "This property is here; it's in this component." In the case of autopoiesis, you can't say that life—the condition of being self-produced—is in this molecule, or in the DNA, or in the cellular membrane, or in the protein. Life is in the configuration and in the dynamical pattern, which is what embodies it as an emergent property.

I find it fascinating to apply this same line of analysis to my own mind, in the cognitive domain. My own sense of self, "me," can be seen in the same light. I have to be relentless to hold on to my identity. These ideas help us to come to a real appreciation of what it means to have an identity—to comprehend what we think of as our own mind. My mind has the quality of "being here" so I can relate to others. For example, I interact; but when I try to grasp it, it's nowhere—it's distributed in the underlying network.

Let me add that this emergence and nonlocality has nothing to do with the current hype about quantum mechanics and the brain. That stuff is perhaps an interesting hypothesis to entertain, but it has no scientific evidence behind it. On the other hand, I'm talking about thirty years' worth of results in cognitive science. I'd go one step further and dispute the typical physicist, who believes that he or she is dealing with fundamental reality. A physicist will say that we're made of atoms. Such statements, while true, are irrelevant. The statement "You're looking at me" doesn't have the same weight as statements concerning the cellular level. There is a reality of life and death, which affects us directly and is on a different level from the abstractions. We have to abandon the enormous deadweight of the materialism of the Western tradition, and turn to a more planetary way of thinking.

##