

## **Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela<sup>1</sup>**

EVAN THOMPSON

*Canada Research Chair in Cognitive Science and the Embodied Mind, Department of  
Philosophy, York University (E-mail: evant@yorku.ca)*

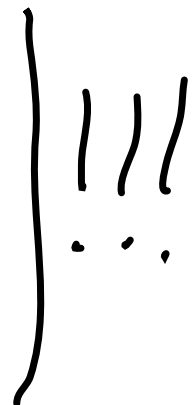
**Abstract.** This talk, delivered at “De l’autopoïèse à la neurophénoménologie: un hommage à Francisco Varela; from autopoiesis to neurophenomenology: a tribute to Francisco Varela,” June 18–20, at the Sorbonne in Paris, explicates several links between Varela’s neurophenomenology and his biological concept of autopoiesis.

**Key words:** autopoiesis, lived body, neurophenomenology, teleology, Varela

Allow me to begin on a personal note. I first met Francisco Varela in the summer of 1977 at a conference called “Mind in Nature.” The conference was organized by my father, William Irwin Thompson, and Gregory Bateson. It took place in Southampton, New York, at the Lindisfarne Association, an institute founded by my father, and was chaired by Bateson, who was then serving as Lindisfarne’s Scholar-in-Residence.<sup>2</sup> I was not quite 15-years old; Francisco was almost 32. At that time Francisco was known within the circle of second-generation cybernetics and systems theory for his work with Maturana on autopoiesis and for his “calculus of self-reference” (Varela et al., 1974; Maturana and Varela, 1975; Varela 1975). But outside this circle he was known for an interview and a paper that had appeared about a year earlier in *Coevolution Quarterly*, a widely read intellectual journal of the American counterculture in the 1970s (Varela 1976a,b). The paper, called “Not One, Not Two,” was a position paper on the mind–body relation, given at another conference, also involving Bateson and Heinz von Foerster, among others. I remember reading this paper, in which Francisco set forth some ideas about dualities and self-reference with application to the mind–body problem, and having the sense that it said something very important, but without being able fully to understand it. I also remember listening to Francisco and the physicist David Finkelstein arguing about the relation between natural systems and logic and mathematics. Francisco was working on the algebraic foundations of self-reference and Finkelstein on quantum logic. Their debate was mesmerizing to me, even though I didn’t have the knowledge or experience to follow it. In preparing for this lecture today, I reread “Not One, Not Two,” not having looked at it carefully for many years. What struck me this time are these words Francisco wrote at the end of the paper: “But what I see as

an important ingredient of our discussion is the fact that a change in experience (being) is as necessary as change in understanding if any suturing the mind–body dualisms is to come about” (Varela 1976a, p. 67). The dualism of concern to Francisco here was not the abstract, metaphysical dualism of mental and physical properties, but rather the dualism of mind as a scientific object versus mind as an experiencing subject. One of the most significant and exceptional aspects of Francisco’s life and work, from this early paper to his last writings on his own illness and liver-transplant experience (Varela, 2001), is that he never lost sight of this point that the mind–body problem is not only a philosophical problem, or a scientific problem, but also a problem of direct experience. The problem could be put this way. It’s one thing to have a scientific representation of the mind as “enactive” – as embodied, emergent, dynamic, and relational; as not homuncular and skull-bound; and thus in a certain sense insubstantial. But it’s another thing to have a corresponding direct experience of this nature of the mind in one’s own first-person case. In more phenomenological terms, it’s one thing to have a scientific representation of the mind as participating in the “constitution” of its intentional objects; it’s another thing to see such constitution at work in one’s own lived experience. Francisco believed, like phenomenologists and also Buddhists, that this kind of direct experience is possible. He also thought that unless science and philosophy make room for this kind of experience, we will never be able to deal effectively with the mind–body problem, but will instead fall prey to one or another extreme view – either denying experience in favour of theoretical constructions, or denying scientific insight in favour of naive and uncritical experience.

Ten years later Francisco and I worked hard on developing these ideas when we began writing our book, *The Embodied Mind*, in 1986 (Varela et al. 1991). If I may be bold, I think that although the ideas about embodied cognition in this book have been widely acknowledged and assimilated by the field, the book’s central theme has yet to be fully absorbed. That theme is the need for back-and-forth circulation between scientific research on the mind and disciplined phenomenologies of lived experience. Without such circulation, the danger for the scientist and philosopher is nihilism, by which I mean the inability to stop experiencing things and believing in them in a way one’s theory says is an illusion. Theoretical ideas like “being no one” (that there are no such things as selves but only neural self-models) (Metzinger, 2003), or that consciousness is the brain’s “user illusion” (Dennett 1991), bear witness to this predicament. An appreciation of what Francisco and I called the “fundamental circularity” of science and experience reminds us that such models of consciousness are objectifications that presuppose, on an empirical level, the particular subjectivities of the scientists who author them, but also, on a transcendental level, the intentionality of consciousness



as an *a priori* openness to reality, by virtue of which we are able to have any comprehension of anything at all. Experience is thus, in a certain sense, irreducible.

Let me jump ahead another 10 years to Francisco's 1996 paper on neurophenomenology (Varela, 1996). Here, the idea that the mind-body problem is also a problem of experience is articulated pragmatically in relation to neuroscience and the so-called "hard problem" of consciousness.

The hard problem of consciousness is the problem of how and why physiological processes give rise to experience. It's one thing to be able to establish correlations between consciousness and brain activity; it's another thing to have an account that explains how and why certain physiological processes suffice for consciousness. At present, we not only lack such an account, but we are also unsure about the form it would need to have in order to overcome the conceptual gap between subjective experience and the brain. In proposing neurophenomenology as a "methodological remedy" for the hard problem, Francisco's insight was that no purely third-person, theoretical proposal or model would suffice to overcome this gap. "In all functionalistic accounts," he wrote, "what is missing is not the coherent nature of the explanation but its alienation from human life. Only putting human life back in will erase that absence; not some 'extra ingredient' or profound 'theoretical fix'" (Varela 1996, p. 345). "Putting human life back in" means, among other things, expanding neuroscience to include original phenomenological investigations of experience. In this way, "the experiential pole enters directly into the formulation of the complete account" (1996, p. 345), rather than being merely the referent of yet another abstract functionalist model. But if experience is to play a central role in this way, then it has to be mobilized according to a rigorous phenomenology. Pragmatically, this means that the neuroscience of consciousness needs to incorporate disciplined, first-person investigations of experience, as illustrated in a preliminary way by one of Francisco's last experimental studies, which used original first-person data to guide the study of brain dynamics (Lutz et al. 2002). Phenomenology is thus "not a convenient stop on our way to a real explanation, but an active participant in its own right" (Varela, 1996, p. 344). In Francisco's words: "disciplined first-person accounts should be an integral element of the validation of a neurobiological proposal, and not merely coincidental or heuristic information" (1996, p. 344).

In addition to this new methodological approach, neurophenomenology is also informed by an autopoietic conception of life, an enactive conception of mind, and a phenomenological conception of intentionality, subjectivity, and the lived body. These link neurophenomenology to what Francisco called "renewed ontologies" of mind and life (Varela, 1997a). This idea of renewed ontologies is what I want to talk about today.

### Life beyond the hard problem

My first step is to recast the terms in which the hard problem of consciousness is usually stated. Consider Thomas Nagel's classic formulation of the hard problem:

If mental processes are physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery (Nagel, 1979, p. 175).

Nagel's point is the now familiar one that we don't understand how an objective physical process could be sufficient for or constitutive of the subjective character of a conscious mental process. But stating the problem this way embeds it within the Cartesian framework of the "mental" versus the "physical," and this framework actually promotes the explanatory gap, and so is incapable of resolving it. **What we need instead is a framework that doesn't set "mental" and "physical" in opposition to each other, or reduce one to the other ("not one, not two"). We need to focus on a kind of phenomenon that is already beyond this gap. *Life* or *living being* is precisely this kind of phenomenon.** For biology, living being is *living organisms*; for phenomenology, it is *living subjectivity*. Where these two meet is in what phenomenologists call the *lived body*. What we need, and what neurophenomenology aims for, is an account of the lived body that integrates biology and phenomenology, and so goes "beyond the gap" (Roy et al., 1999).

What happens if we substitute "body" for "physical" in Nagel's statement?

If mental processes are bodily processes, then there is something it is like, intrinsically, to undergo certain bodily processes.

Does this substitution make any difference? If there belongs to certain bodily processes something it is like to undergo them, then those bodily processes are experiences. They have a subjective or first-person character, which they could not lack without ceasing to be experiences. **They are *feelings*,** in the broad sense William James had in mind when he used the word "feeling" "to designate all states of consciousness merely as such" (James, 1981, p. 185), and that Damasio has revived by describing **feelings as "bearing witness to life within our minds"** (Damasio, 2003, p. 140). The problem of what it is for mental processes to be also bodily processes is thus in large part the problem of *what it is for subjectivity and feeling to be a bodily phenomenon*. In phenomenological language:

What is it for a physical living body (*Körper/leiblicher Körper*) to be also a lived body (*Leib/körperlicher Leib*)?

It's tempting to call this problem the *body–body problem*. I offer it as a “radical embodiment” reformulation of the hard problem (Thompson and Varela, 2001; Thompson, forthcoming)

In putting the problem this way, I am relying on the phenomenological distinction between the body as a material thing (*Körper*), and the body as a living and feeling being (*Leib*). This distinction is between two modes of appearance of one and the same body, not between two bodies or two properties (in the property-dualist sense). Hence the explanatory gap is now between two types within one typology of embodiment or living being, not between two opposed and reified ontologies (“mental” and “physical”). Furthermore, this gap is no longer absolute, because in order to state it we need to make common reference on both sides to *life* or *living being*.

What is the difference between two modes of appearance and property dualism?

These two points are philosophically nontrivial. In the hard problem as classically conceived, the gap is absolute, because there is and can be no conceptual unity to the mental and the physical, consciousness and the brain. Consciousness is equated with qualia, which are supposed to be phenomenal properties that resist functional analysis, while the body is equated with structure and function, with mechanism.<sup>3</sup> Given these equivalences, one must either mechanize consciousness in order to reduce it to a brain state, or be a property dualist. This way of dividing up the universe is thoroughly Cartesian. Although physicalist philosophy of mind today rejects Descartes's substance dualism, it maintains both the underlying conceptual separation of mind and life, and the equation of life with mere mechanism.<sup>4</sup>

Yes!

For neurophenomenology, by contrast, the guiding issue isn't the contrived problem of how to derive a subjectivist concept of consciousness from an objectivist concept of the body. Instead, it's to understand the *emergence of living subjectivity from living being, including the reciprocal shaping of living being by living subjectivity*. It's this issue of *emergence* that neurophenomenology addresses, not the Cartesian version of the hard problem.

Don't understand this

### The strong continuity of life and mind

Implicit in this step of recasting the terms of the hard problem is the idea of a *strong continuity* of life and mind. One way to put this idea is that life and mind share a common pattern or organization, and the organizational properties characteristic of mind are an enriched version of those fundamental to life (Godfrey-Smith, 1996, p. 230; see also Wheeler 1997). Mind is life-like, and life is mind-like. But a simpler and more provocative formulation is this one: ***Living is cognition.***

This proposition comes from Maturana and Varela's theory of autopoiesis (Maturana and Varela, 1980). Some have taken the “is” in this proposition as the “is” of identity (living = cognition) (Stewart 1992, 1996), others as

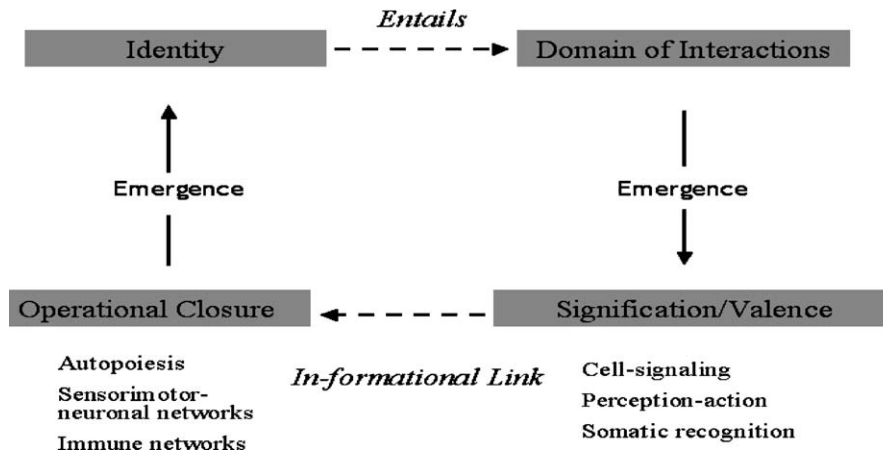
the “is” of predication or class inclusion (all life is cognitive) (Bourguine and Stewart, in press; Bitbol and Luisi, forthcoming). The origins of the proposition go back to Maturana’s 1970 paper, “Biology of Cognition” (Maturana 1970). There he used the concept of cognition widely to mean the operation of any living system in the domain of interactions specified by its circular and self-referential organization. Cognition is effective conduct in this domain of interactions, not the representation of an independent environment. In Maturana’s words: “*Living systems are cognitive systems, and living as a process is a process of cognition. This statement is valid for all organisms, with and without a nervous system*” (Maturana 1970 p. 13).

This could answer the boundary problem of Henry’s Life

Francisco later came to prefer a different way of explicating the “living is cognition” proposition: *Living is sense-making*. Consider motile bacteria swimming uphill in a food gradient of sugar. The cells tumble about until they hit on an orientation that increases their exposure to sugar, at which point they swim forward, up-gradient, toward the zone of greatest sugar concentration. This behaviour happens because the bacteria are able to sense chemically the concentration of sugar in their local environment through molecular receptors in their membranes, and they are able to move forward by rotating their flagella in coordination like a propeller. These bacteria are of course autopoietic. They also embody a dynamic sensorimotor loop: the way they move (tumbling or swimming forward) depends on what they sense, and what they sense depends on how they move. Moreover, the sensorimotor loop both expresses and is subordinated to the system’s autonomy, to the maintenance of its autopoiesis. As a result, every sensorimotor interaction and every discriminable feature of the environment embodies or reflects the bacterial perspective. For instance, *although sucrose is a real and present condition of the physicochemical environment, its status as food is not*. That sucrose is a nutrient isn’t intrinsic to the structure of the sucrose molecule; it’s a relational feature, linked to the bacterium’s metabolism. Sucrose has significance or value as food, but only in the milieu that the organism itself brings into existence. Francisco summarized this idea by saying that thanks to the organism’s autonomy, its world or niche has a “surplus of significance” compared with the physicochemical environment (Varela, 1991 1997b). Living isn’t simply a cognitive process; it’s also an *emotive* process of sense-making, of bringing signification and value into existence. In this way the world becomes a place of *valence*, of attraction and repulsion, approach or escape. This idea can be depicted in the diagram below (Varela 1997b):

Using this representation, I would like to expand the proposition “living is sense-making” in the following way:

1. *Life = autopoiesis*. By this I mean the thesis that the three criteria of autopoiesis – (i) a boundary, containing (ii) a molecular reaction network,



that (iii) produces and regenerates itself and the boundary – are necessary and sufficient for the organization of minimal life.

2. *Autopoiesis entails emergence of a self.* A physical autopoietic system, by virtue of its operational closure, gives rise to an individual or self in the form of a living body, an organism.
3. *Emergence of a self entails emergence of a world.* The emergence of a self is also by necessity the emergence of a correlative domain of interactions proper to that self, an *Umwelt*.
4. *Emergence of self and world = sense-making.* The organism's world is the sense it makes of the environment. This world is a place of significance and valence, as a result of the global action of the organism.
5. *Sense-making = cognition (perception/action).* Sense-making is tantamount to cognition, in the minimal sense of viable sensorimotor conduct. Such conduct is oriented toward and subject to signification and valence. Signification and valence do not pre-exist "out there," but are enacted or constituted by the living being. Living entails sense-making, which equals cognition.

At this point you may want to object that the proposition "life is cognition" conflates *cognition* with *adaptation*. Margaret Boden (2000, p. 40) makes this charge. She thinks it would be better to use the term "cognition" more strictly to avoid the implication that autopoiesis necessarily involves cognition.

[Read this](#)

I disagree. We need to ask what exactly is meant by "adaptation." For Neo-Darwinians evolution involves the *optimization of adaptation* through natural selection. But from the autopoietic perspective, evolution involves simply the *conservation of adaptation*: as long as a living being doesn't disintegrate, but maintains its autopoietic integrity, it is *ipso facto* adapted, because its mode of sense-making continues to be viable. From this point of view, adaptation is an

invariant background condition of all life. “Cognition,” on the other hand, in the present context means the sense-making activity of living, which underlies the conservation of adaptation – no sense-making, no living, no conservation of adaptation. Notice that this way of thinking about cognition rests on an explicit hypothesis about the natural roots of intentionality: *intentionality arises from the operational closure of an autonomous system, whose paradigm and minimal case is an autopoietic system*. This hypothesis also amounts to a proposal about how to connect the phenomenological conception of intentionality to biology and complex systems theory.

At the other end of the spectrum from Boden, the biologist Lynn Margulis speaks of “microbial consciousness” and suggests that the “conscious cell” is the evolutionary antecedent of animal consciousness and the nervous system (Margulis, 2001). And the phenomenologist Maxine Sheets-Johnstone, in her book *The Primacy of Movement*, argues that bacteria aren’t simply cognitive, but embody a rudimentary kind of corporeal consciousness (Sheets-Johnstone, 1999, pp. 52, 73).

Other refs

You might be tempted to dismiss this idea of cellular consciousness out of hand. But let’s consider the idea for a moment. “Consciousness” can have many meanings, but the one most relevant here is *sentience*, the feeling of being alive and exercising effort in movement. Maine de Biran wrote of *le sentiment de l’existence*. Damasio (1999) and Panksepp (1998) talk about a primitive *feeling of self*. Phenomenologists, from Patocka (1998) to Sheets-Johnstone (1999) to Barbaras (1999), call attention to the importance of *movement* for understanding the intentionality of consciousness. Margulis (2001), like Rodolfo Llinás (2001), describes conscious thought as mental movement. She believes that as brain activity it derives from ancient motile bacteria, which have left their evolutionary stamp on the cellular architecture and communication of neurons. One might summarize these threads by saying that consciousness as sentience is a kind of *primitively self-aware liveliness or animation of the body*. Does this emerge with life itself, with the very first living bodies, namely, bacterial cells? Hans Jonas poses the problem clearly:

At which point . . . in the enormous spectrum of life are we justified in drawing a line, attributing a “zero” of inwardness to the far side and an initial “one” to the side nearer to us? *Where else but at the very beginning of life can the beginning of inwardness be located?* (Jonas, 1996, p. 63; see also Jonas, 1966, pp. 57, 58).

Whether we give this inwardness the name of feeling, receptiveness or response to stimuli, volition, or something else – it harbors, in some degree of “awareness,” the absolute interest of the organism in its own being and continuation (Jonas, 1996, p. 69; see also Jonas, 1966, p. 84).

This “absolute interest of the organism in its own being and continuation” is what Spinoza called *conatus*, the “concern” to exist, to carry on being, that



belongs to life. Jonas observes that Spinoza, with the knowledge of his time, didn't realize that this concern "can only operate as a movement that goes constantly *beyond* the given state of things" (Jonas, 1968, p. 243), and so is never a matter of mere preservation. We, however, can observe that Jonas, with the knowledge of his time, didn't realize that this self-transcending movement is a natural consequence of a certain kind of self-organizing, morphodynamic system, namely, an autopoietic one. The self-transcending movement of life is none other than metabolism, and metabolism is none other than the biochemical instantiation of the autopoietic organization. That organization must remain invariant, otherwise the organism dies, but the only way autopoiesis can stay in place is through the incessant material flux of metabolism. In other words, the operational *closure* of autopoiesis demands that the organism be an *open system*. Jonas called this condition the "needful freedom" of the organism. The organism is never bound to its material composition at any given instant, but by the same token it has to change, because stasis means death.

Coming back to the question about consciousness, I think that life's sense-making is a manifestation of the organism's autonomy and coupling, but not necessarily of consciousness. In support of my preference for this view I would appeal to the following considerations. First, being "phenomenally conscious" of something would seem to entail being able to form intentions to act in relation to it (Hurley 1998, pp. 149–150). It's hard to make sense of the idea of being conscious of something, in the sense of subjectively experiencing it, while having no intentional access to it whatsoever. But there seems no reason to think that autopoietic selfhood of the minimal cellular sort involves any kind of intentional access on the part of the organism to its sense-making. Second, it seems unlikely that minimal autopoietic selfhood involves phenomenal selfhood or subjectivity, in the phenomenological sense of a pre-reflective self-awareness constitutive of a phenomenal first-person perspective (Zahavi 1999). Rather, this would seem to require the reflexive elaboration and interpretation of life processes provided by the nervous system. Finally, it's important to situate consciousness in relation to dynamic, unconscious processes of life regulation, and this becomes difficult if one projects consciousness down to the cellular level.

### Teleology and "autopoietic machines"

A number of things I've said so far suggest that living beings are in some sense teleological: organisms have an interest in their own being and continuation; they realize a dynamic impulse to carry on being; they are always impelled beyond their present condition – these are teleological modes of description. "Living is sense-making" also sounds like a teleological description, because it characterizes the organism as oriented toward the sense it makes of its environment. "Sense-making" is reminiscent of the phenomenological notion

of intentionality, which signifies not a static representational “aboutness,” but rather an act of intending, a purposive striving focused on finding satisfaction in further cognitive acquisitions and experience (see Held, 2003 p. 14). Behind this concept of intentionality we can see the metaphor or kinaesthetic image schema (in Lakoff and Johnson’s sense) of self-generated and goal-directed movement, the motility of life.

Yet how are we to understand this suggestion of teleology in relation to the theory of autopoiesis, which in its original formulation was mechanistic and anti-teleological? Maturana and Varela explicitly identified living systems with machines, and denied that living systems are teleological: “Living systems, as physical autopoietic machines, are purposeless systems” (Maturana and Varela, 1980, p. 86). By “machine” they clearly did not mean an artifact. They meant any system whose operation is determined by its relational organization and the way that organization is structurally realized.<sup>5</sup> Autopoietic systems maintain their own organization constant through material change, and thus are homeostatic (or homeodynamic) systems of a special sort (Maturana and Varela 1980, pp. 78, 79).

At this point we need to ask whether having a relational organization is sufficient for being a machine. We can also wonder about the notion of emergence in this context. The work Francisco and I did on emergence and whole-system causation would seem to conflict with his view that autopoiesis can be realized in a cellular automaton (Thompson and Varela, 2001). In a cellular automaton, there is arguably no genuine emergence and system causation, because every unit is local and the global pattern is in the eye of the observer. We argued, however, that in real living systems, such as a cell or large-scale neural assembly, there is emergence and circular causality, such that the system moves as a whole and constrains the states of its components.

At this juncture, I think it may be useful to draw from another line of work in theoretical biology, the work of Robert Rosen (1991, 2000). Rosen and Francisco share many ideas, although oddly they never mention each other in their writings.<sup>6</sup> Rosen’s dictum is that organisms are different from machines because they are “closed to efficient causation” (Rosen 1991, p. 244). In an organism, but not in a machine, every efficient cause is produced inside the organism. More abstractly stated, Rosen argues that in a relational model of an organism every function (in the mathematical sense of a mapping) is entailed by another function within the model, whereas in a relational model of a machine this closure doesn’t obtain, and one has to go outside the system and appeal to the environment. As Rosen (1991, p. 246) puts it, there’s an “impoverishment of entailment” in a machine compared with an organism. In Francisco’s language, this difference corresponds to the difference between an autonomous system with operational closure and a heteronomous system defined by outside control (Varela 1979). But Rosen also argues that closure

and maximal entailment in an organism can't be simulated by a Turing machine (Rosen, 1991, 2000, pp. 266–269). More precisely, he shows that a certain class of relational models called *Metabolism-Repair* systems or  $(M,R)$  systems, in which every function is entailed by another function inside the system, aren't Turing-computable. On this basis, he argues that any material realization of an  $(M,R)$  system, such as a cell, can't be a mechanism or machine. This raises the question of what the relation is between Rosen's  $(M,R)$  systems and autopoietic systems. In a recent article, Letelier et al. (2003) argue that *autopoietic systems are a subset of Rosen's  $(M,R)$  systems*: every autopoietic system is operationally equivalent to an  $(M,R)$  system (but not conversely, because a generic  $(M,R)$  system lacks the autopoietic property of generating its own boundary and internal topology). It would seem to follow that *autopoietic systems are not Turing-computable*, and that a physical autopoietic system – an organism or living being – is not a machine (at least according to one abstract and powerful concept of mechanism).

I have to understand better what non Turing-computability means

If Rosen is right about life being noncomputable, then this result is an important challenge to the original placement of autopoiesis in the category of cybernetic mechanism. It also challenges the hypothesis that autopoiesis can be captured by cellular automata models, and allows for a stronger notion of emergence than the emergence we see in cellular automata. Emergence is present when there is no way to analyze a system into pre-existing parts and resultant whole. Maturana and Varela, and Rosen, in different ways both argued that this sort of analysis or “fractionation” fails in the face of the organism's self-referential organization. Here, part and whole are completely interdependent: an emergent whole is produced by a continuous interaction of its parts, but these parts cannot be characterized independently from the whole.

We can now return to the issue of teleology. Francisco, in his articles up to the early 1990s, continued to resist the idea that autopoiesis involves anything teleological (see Varela, 1991, 1997b). But in one of his last essays he changed his mind. This essay, written with Andreas Weber (Weber and Varela, 2002), concerns autopoiesis and the problem of teleology and the organism from Kant's *Critique of Judgement*. There Francisco argues that *teleology arises from autopoiesis* and is none other than the organism's sense-making. Yet strangely he doesn't even mention, let alone discuss, the change from his earlier to later view, nor the reasons for the change. Nor does he comment on his earlier acceptance of the life-as-machine notion – another striking omission considering that one of Kant's main points was that organisms are “natural purposes” by virtue of being self-organizing, and must be judged to be fundamentally different from machines.

Francisco's change of view reflects his immersion in phenomenology at the end of his life. Even his later articles disavowing teleology were written prior

to this neurophenomenological phase.<sup>7</sup> The change of view also reflects his deeper study of traditions of biological thought influenced by Kant. The type of teleology Maturana and Varela criticized in *Autopoiesis and Cognition* was *teleonomy* or Neo-Darwinian functionalistic explanation. But the type of teleology Francisco later discussed is Kant's idea that the organism is a "natural purpose" because it is a self-organizing being (Kant, 1951, §64–65). Francisco came to think that autopoiesis provides a naturalistic way of grounding, reformulating, and advancing the Kantian view of living beings as teleological, in a way Kant thought was impossible.<sup>8</sup>

Francisco never tried to reconcile his earlier and later statements about teleology, but I think it might be possible to do so in the following way. The main point he and Maturana insisted upon in *Autopoiesis and Cognition* is that teleology does not belong to the *autopoietic organization*. This point remains valid: in setting out the conditions for a self-producing organization in the molecular domain, no reference is made to notions such as "end," "purpose," "goal," or "function." On the other hand, the main point of the later revision is that teleology is none other than sense-making. Sense-making is not a feature of the autopoietic *organization*, but rather of the *coupling* of a concrete autopoietic system and its environment. In other words, teleology is not an intrinsic organizational property, but an emergent relational one that belongs to a concrete autopoietic system interacting with its environment.

Let me try to indicate where these reflections seem to lead. If living beings are not reducible to algorithmic mechanism, and if teleology is an emergent relational property, not an intrinsic organizational one, then we are faced with the prospect of a new kind of biological naturalism beyond the classical opposition of mechanism and teleology. Francisco's intuition was that such a naturalism would be able to offer strong bridges to phenomenology, but also that phenomenology could contribute to its formulation. Thus naturalizing phenomenology, for Francisco, always implied a corresponding phenomenological reconceptualization of nature.

In preparing this talk, I was struck by the thought that maybe one reason Francisco revised his view about teleology, though perhaps not a consciously articulated one, was his immersion in the life process of his own chronic and terminal illness. Francisco experienced first-hand, in an intense and singular way, life's sense-making. He realized it through his experience of his own living being, as it suffered the changing anti-viral treatments, the liver transplant and its "offering" of life,<sup>9</sup> the chemotherapy, the fatigue of sickness, and his scientific and phenomenological curiosity about living and dying. Using a Freudian idiom, Francisco called this curiosity his "epistemo-philia." I would add that his epistemo-philia was unique, in its embodiment of Buddhist mental presence, mathematical insight, phenomenological intuition, and an exceptional biologist's "feeling for the organism."<sup>10</sup> Francisco's revisiting the

problem of teleology reflects his deep insight that *the mind–body problem is first and foremost a problem of lived experience*. As he and Weber wrote in this last article on life and teleology, commenting on Kant and Jonas:

It is actually by experience of our teleology – our wish to exist further on as a subject, not our imputation of purposes on objects – that teleology becomes a real rather than an intellectual principle. . . before being scientists we are first living beings, and as such we have the evidence of intrinsic teleology in us. And, in observing other creatures struggling to continue their existence – starting from simple bacteria that actively swim away from a chemical repellent – we can, by our own evidence, understand teleology as the governing force of the realm of the living. Theories about the living can only be conceived from the fragile and concerned perspective of the living itself: [and then, quoting Jonas] “. . . *life can only be known by life*” (Weber and Varela 2002 p. 110).

### **Life can be known only by life**

To close this talk, I would like to comment on this proposition that life can be known only by life. The claim is a transcendental one in a Kantian and Husserlian sense: it’s about the conditions for the possibility of knowing life, given that we do actually have biological knowledge. Consider the question, how is it that we are able to recognize or comprehend the form or dynamic pattern of autopoiesis in the first place? Would this pattern be recognizable at all from some ideal objective standpoint? Or is it rather that we’re able to recognize this pattern only because it resembles the form of our own bodily selfhood, which we know first-hand? Here, in brief, is the phenomenologist’s answer: (1) An adequate account of certain observable phenomena requires the concepts *organism* (in the original Kantian sense of a self-organizing whole) and *autopoiesis*. (2) The source of the meaning of these concepts is the *lived body* – our first-person, lived experience of our own animate, bodily existence. (3) These concepts and the biological accounts in which they figure aren’t derivable, even in principle, from some observer-independent, nonindexical, objective, physico-functional description (according to the physicalist myth of science). As Jonas puts it, no disembodied and purely intellectual mind, like Laplace’s divine mathematician, would be able to comprehend the *form* of the organism simply from a complete knowledge of the microphysical state of things. To make the link from matter to life and mind, from physics to biology, one needs concepts like *organism* and *autopoiesis*, but such concepts are available only to an embodied mind with first-hand experience of its own living body. In Merleau-Ponty’s words: “*Je ne puis comprendre la fonction du corps vivant qu’en l’accomplissant moi-même et dans la mesure où je suis un corps qui se lève vers le monde.*” (“I cannot understand the function of the living body except by enacting it myself, and except in so far as I am a body which rises toward the world.”)<sup>11</sup>

Francisco said that the “basic ground” of neurophenomenology is the “irreducible nature of conscious experience.” “Lived experience,” he wrote, is “where we start from and where we all must link back to, like a guiding thread” (Varela 1996, p. 334). Let us be clear about what this means. Experience is irreducible not because it possesses metaphysically peculiar “properties” that can’t be squeezed into some reified, physicalist model of the universe, after the fashion of contemporary property dualism. It’s irreducible because of its ineliminable transcendental character: lived experience is always already presupposed by any statement, model, or theory, and the lived body is an *a priori* invariant of lived experience. Experience is *die unhintergehbarkeit* – the “ungobehindable.” There is no dualism or idealism here: **the transcendental lived body is no other than the empirical living body; it’s simply that body re-membered in a certain way** – namely, as where we start from and where we must all link back to, like a guiding thread.

I began this talk on a personal note, and I would like to end it that way too. The first conversation I ever had with Francisco was while we drove together with my father from New York City to Southampton to the “Mind in Nature” conference in 1977. Not long before I had discovered the writings of Borges, which I proceeded to devour in the way only an adolescent mind can. Somehow Francisco and I fell into a conversation about literature, and I declared my enthusiasm for Borges. Francisco preferred Neruda. About a year later Francisco gave me a copy of the English translation of Neruda’s *Memoirs*, inscribed, “To Evan Thompson, with love and friendship, Francisco, September 1978,” which I still have to this day. On the first page of this book, Neruda writes: “Perhaps I didn’t live just in my self, perhaps I lived the lives of others. . . . My life is a life put together from all those lives: the lives of the poet.” These words express a sentiment that Francisco’s thought and life echo in so many ways. Speaking for myself, my talk today is an expression of my deep gratitude for the participation of Francisco’s life in mine.

## Notes

1. This text was presented to the conference “De l’autopoïèse à la neurophénoménologie: un hommage à Francisco Varela; from autopoiesis to neurophenomenology: A tribute to Francisco Varela,” June 18–20, at the Sorbonne in Paris. The text was written to be read aloud, and I have resisted the urge to alter it, so that it can remain true to its inspiration and purpose.
2. The conference ran from August 24–31, 1977. The participants were Lewis Balamuth, Gregory Bateson, Mary Catherine Bateson, David Finkelstein, David Fox, William Irwin Thompson, Francisco Varela, and Arthur Young. It took place while Bateson was working on the manuscript of his last book, *Mind and Nature: A Necessary Unity* (New York: E.P. Dutton, 1979).
3. See Chalmers (1996, 1997) and Kim (1998).

4. This is clearly evident, for instance, in the widespread view that there is no hard problem of life, because life is nothing but structure and function, whereas there is for consciousness, because physical structure and function logically underdetermine phenomenal consciousness. See Chalmers (1996, pp. 106–107, 169; 1997, pp. 5–6).
5. Maturana and Varela (1980, pp. 75, 77). Varela (1979, p. 7) says in a footnote: “In this book ‘machines’ and ‘systems’ are used interchangeably. They obviously carry different connotations, but the differences are inessential, for my purpose, except in seeing the relation between the history of biological mechanism and the modern tendency for systemic analysis. Machines *and* systems point to the characterization of a class of unities in terms of their organization.”
6. Like Francisco, Rosen died not long ago (in 1998) and too young (64), from complications of an illness. He lived in my home country, Canada, and I regret that I didn’t study his work in time to talk to him and Francisco about their ideas. A worthwhile scientific and epistemological project would be to assess their theories in relation to one another, and I’m happy to see that the Chilean team of Letelier et al. (2003) has begun this work.
7. Varela (1991) and (1997b). The second article was written for a conference in 1992.
8. Recently I rediscovered some email correspondence Francisco and I had in June 1999 about this issue of teleology. It began because I pointed out to him that his commitment to phenomenology seemed inconsistent with his older position on teleology with Maturana. We had both independently been reading Kant and Jonas, and I asked Francisco whether he would still maintain his earlier anti-teleological stance in light of Jonas’s argument that one cannot recognize something to be a living being unless one recognizes it as purposive, and that one cannot recognize something as purposive unless one is an embodied agent who experiences purposiveness in one’s own case. Francisco replied that he was “still quite suspicious” about this appeal to teleology, and hence this way of linking phenomenology and biology, and that he preferred to “shift the accent” from teleology to original intentionality, understood as the sense-making capacity proper to autopoietic units. He saw this shift as a refinement of the “‘Santiago school’ move to introduce the equation life = cognition.” It’s clearly “silly,” he said, to make cellular cognition just like animal cognition, but their “common root” is this basic sense-making capacity proper to autopoietic life. Appealing to sense-making, he suggested, was more “constructive” than appealing to the “elusive principle of purpose.” Sense-making provides a strong link to intentionality, but “whether this turns into teleology,” he said, “is another matter.” This line of thought, however, struck me as unsatisfactory, because “original intentionality” and “sense-making” are themselves arguably teleological notions. The issue is precisely how to analyze this teleology. So although the proposition “living is sense-making” may be an important elaboration of the equation life = cognition, it is insufficient to establish the anti-teleological stance with respect to Kant’s and Jonas’s notion of teleology. Six months later in December 1999, in response to another email of mine pressing him on this issue, Francisco indicated that as time had gone by he had come to have a “broader view,” and to see that “in a funny way you do recover a full fledged teleology. . . but this teleology is. . . intrinsic to life in action,” and “does not require an extra transcendental source” in the Kantian sense. In other words, teleology, in the sense of self-organized natural purposiveness, can be seen as an empirical feature of the organism, based on its autonomy and sense-making, rather than only a form of judgement, as Kant had held. It’s precisely this conception that Weber and Varela (2002) presented, and that they called *intrinsic teleology*.
9. See Varela (2001).
10. I borrow this phrase from Keller (1984).
11. French: Merleau-Ponty (1945, p. 90). English: Merleau-Ponty (1962, p. 75).

## References

- Barbaras, R. 1999. The movement of the living as the originary foundation of perceptual intentionality. In: Petitot, Varela, Pachoud, and Roy (eds), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, pp. 525–538. Stanford, CA: Stanford University Press.
- Bitbol, M. and Luisi, P. L. forthcoming. Autopoiesis with or without cognition: Defining life at its edge.
- Boden, M. 2000. Autopoiesis and life. *Cognitive Science Quarterly* 1, 117–145.
- Bourgine, P. and Stewart, J. in press. Autopoiesis and cognition. *Artificial Life*.
- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Chalmers, D. J. 1997. Moving forward on the problem of consciousness. *Journal of Consciousness Studies* 4, 3–46.
- Damasio, A. R. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.
- Damasio, A. 2003. *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. New York: Harcourt.
- Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little Brown.
- Godfrey-Smith, P. 1996. Spencer and Dewey on life and mind. In: M. Boden (ed), *The Philosophy of Artificial Life*, pp. 314–331. Oxford: Oxford University Press.
- Held, K. 2003. Husserl's phenomenological method. (translated by L. Rodemeyr). In: D. Welton (ed), *The New Husserl: A Critical Reader*, pp. 3–31. Bloomington and Indianapolis: Indiana University Press.
- Hurley, S. L. 1998. *Consciousness in Action*. Cambridge, MA: Harvard University Press.
- James, W. 1981. *The Principles of Psychology*. Cambridge, MA: Harvard University Press.
- Jonas, H. 1966. *The Phenomenon of Life: Toward a Philosophical Biology*. Chicago: University of Chicago Press.
- Jonas, H. 1968. Biological foundations of individuality. *International Philosophical Quarterly* 8, 231–251.
- Jonas, H. 1996. *Mortality and Morality: A Search for the Good After Auschwitz*. Evanston, IL: Northwestern University Press.
- Kant, I. 1951. *Critique of Judgement* (translated by J. H. Bernard). New York: Hafner Press.
- Keller, E. F. 1984. *A Feeling for the Organism: The Life and Work of Barbara McClintock*. New York: W.H. Freeman.
- Kim, J. 1998. *Mind in a Physical World. An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: The MIT Press/A Bradford Book.
- Letelier, J. C., Marín, G. and Mpodozis, J. 2003. Autopoietic and (M,R) systems. *Journal of Theoretical Biology* 222, 261–272.
- Llinás, R. 2001. *I of the Vortex: From Neurons to Self*. Cambridge, MA: MIT Press.
- Lutz, A., Lachaux, J.-P., Martinerie, J. and Varela, F. J. 2002. Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Sciences USA* 99, 1586–1591.
- Margulis, L. 2001. The conscious cell. In: P. C. Marijúan (ed), *Cajal and Consciousness: Scientific Approaches to Consciousness on the Centennial of Ramon y Cajal's Textura. Annals of the New York Academy of Sciences*, Vol. 929, pp. 55–70. New York: New York Academy of Sciences.
- Maturana, H. R. 1970. Biology of cognition. In: Maturana and Varela 1980.



- Maturana, H. R. and Varela, F. J. 1975. Autopoietic systems: A characterization of the living organization. Biological Computer Lab Report 9.4, University of Illinois, Urbana. Reprinted in Maturana and Varela, 1980.
- Maturana, H. R. and Varela, F. J. 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy of Science, Vol. 42. Dordrecht: D. Reidel.
- Merleau-Ponty, M. 1945. *Phénoménologie de la perception*. Paris: Gallimard.
- Merleau-Ponty, M. 1962. *Phenomenology of Perception* (translated by Colin Smith). London: Routledge Press.
- Metzinger, T. 2003. *Being No One. The Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.
- Nagel, T. 1979. What is it like to be a bat? Reprinted in T. Nagel, *Mortal Questions*, pp. 165–180. New York: Cambridge University Press.
- Panksepp, J. 1998. The periconscious substrates of consciousness: Affective states and the evolutionary origins of self. *Journal of Consciousness Studies* 5, 566–582.
- Patocka, J. 1998. *Body, Community, Language, World* (translated by E. Kohák). Chicago and La Salle: Open Court.
- Rosen, R. 1991. *Life Itself. A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. New York: Columbia University Press.
- Rosen, R. 2000. *Essays on Life Itself*. New York: Columbia University Press.
- Roy, J.-M., Petitot, J., Pachoud, B. and Varela, F. J. 1999. Beyond the gap: An introduction to naturalizing phenomenology. In: J. Peitot, F. J. Varela, B. Pachoud, and J.-M. Roy (eds), *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, pp. 1–80. Stanford, CA: Stanford University Press.
- Sheets-Johnstone, M. 1999. *The Primacy of Movement*. Amsterdam and Philadelphia, PA: John Benjamins Press.
- Stewart, J. 1992. Life = cognition: The epistemological and ontological significance of artificial life. In: P. Bourgin and F. J. Varela (eds), *Toward a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life*, pp. 475–483. Cambridge, MA: MIT Press.
- Stewart, J. 1996. Cognition = life. Implications for higher-level cognition. *Behavioural Processes* 35, 311–326.
- Thompson, E. forthcoming. *Radical Embodiment: The Lived Body in Biology, Human Experience, and the Sciences of Mind*. Harvard University Press.
- Thompson, E. and Varela, F. J. 2001. Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences* 5, 418–425.
- Varela, F. J. 1975. A calculus for self-reference. *International Journal of General Systems* 2, 5–24.
- Varela, F. J. 1976a. On observing natural systems (interview by Donna Johnson). *CoEvolution Quarterly*, 11(Summer 1976), 26–31.
- Varela, F. J. 1976b. Not one, not two. *Coevolution Quarterly*, 12(Fall 1976), 62–67.
- Varela, F. J. 1979. *Principles of Biological Autonomy*. New York: Elsevier North Holland.
- Varela, F. J. 1991. Organism: A meshwork of selfless selves. In: A. Tauber (ed), *Organism and the Origin of Self*, pp. 79–107. Dordrecht: Kluwer Academic Publishers.
- Varela, F. J. 1996. Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies* 3, 330–350.
- Varela, F. J. 1997a. The naturalization of phenomenology as the transcendence of nature: Searching for generative mutual constraints. *Alter* 5, 355–381.
- Varela, F. J. 1997b. Patterns of life: Intertwining identity and cognition. *Brain and Cognition* 34, 72–87.

- Varela, F. J. 2001. Intimate distances: Fragments for a phenomenology of organ transplantation. In: E. Thompson (ed), *Between Ourselves: Second-Person Issues in the Study of Consciousness*, pp. 259–271. Thorverton, UK: Imprint Academic. Also in *Journal of Consciousness Studies* 8, 259–271.
- Varela, F. J. and Depraz, N. 2000. At the source of time: Valence and the constitutional dynamics of affect. *Arob@se. Journal de lettre et de sciences humain* 4(1–2). <http://www.arobase.to>. Also published in S. Gallagher and S. Watson (eds), *Ipseity and Alterity: Interdisciplinary Approaches to Intersubjectivity*. Rouen: Presses Universitaires de Rouen, in press.
- Varela, F. J., Maturana, H. R., and Uribe, R. 1974. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems* 5, 187–196.
- Varela, F. J., Thompson, E., and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.
- Weber, A. and Varela, F. J. 2002. Life after Kant: Natural purposes and the autopoietic foundations of individuality. *Phenomenology and the Cognitive Sciences* 2, 97–125.
- Wheeler, M. 1997. Cognition's coming home: The reunion of life and mind. In: P. Husbands and I. Harvey (eds), *Proceedings of the 4th European Conference on Artificial Life*, pp. 10–19. Cambridge, MA: The MIT Press.
- Zahavi, D. 1999. *Self-Awareness and Alterity. A Phenomenological Investigation*. Evanston, IL: Northwestern University Press.