Essay 1

Autopoiesis and a Biology of Intentionality*

Francisco J. Varela

CREA, CNRS—Ecole Polytechnique,
Paris, France.

Parts of this text have been published in (Varela 1991).

1.1 Introduction

As everybody here knows, autopoiesis is a neologism, introduced in 1971 by H. Maturana and myself to designate the organization of a minimal living system. The term became emblematic of a view of the relation between an organism and its medium, where its self constituting and autonomous aspects are put at the center of the stage. From 1971, until now much has happened to reinforce this perspective. Some of the developments have to do with the notion of autopoiesis itself in relation to the cellular organization and the origin of life. Much more has to do with the autonomy and self-organizing qualities of the organism in relation with its cognitive activity. Thus in contrast to the dominant cognitivist, symbol-processing views of the 70's today we witness in cognitive science a renaissance of the concern for the embeddedness of the cognitive agent, natural or artificial. This comes up in various labels as nouvelle-AI (Brooks 1991c), the symbol grounding problem (Harnad 1991), autonomous agents in artificial life (Varela & Bourgine 1992), or situated functionality (Agree 1988), to cite just a few self-explanatory labels used recently.

Any of these developments could merit a full talk; obviously I cannot do that here. My intention rather, profiting from the position of opening this gathering, is to try to indicate some fundamental or foundational issues of the relation between autopoiesis and perception. Whence the title of my talk: a biology of intentionality. Since the crisis of classical cognitive science has thrown open the issue of intentionality, in my eyes autopoiesis provides a natural entry into a view of intentionalty that is seminal in answering the major obstacles that have been addressed recently. I'll came back to that at the end. Let me begin at the beginning.

1.2 Cognition and Minimal Living Systems

1.2.1 Autopoiesis as the skeletal bio-logic

The bacterial cell is the simplest of living systems because it possesses the capacity to produce, through a network of chemical processes, all the chemical components which lead to the constitution of a distinct, bounded unit. To avoid being trivial, the attribute 'living' in the foregoing description must address the *process* that allows such constitution, not the materialities that go into it, or an enumeration of properties. But what is this basic process? Its description must be situated at a very specific level: it must be sufficiently universal to allow us to recognize living systems as a class, without essential reference to the material components. Yet at the same time it must not be too abstract, that is, it must be explicit enough to allow us to see such dynamical patterns in action in the actual living system we know on earth, those potentially to be found in other solar systems, and eventually those created artificially by man. As stated by the organizer of a meeting on artificial life: "Only when we are able to view *life-as-we-know-it* in the larger context of *life-as-it-could-be* will we really understand the nature of the beast" (Langton 1989b, p. 2).

Contemporary cell biology has made it possible for some years now to put forth the characterization of this basic living organization—a bio-logic—as that of an *autopoietic* system (from Greek: self-producing—Maturana & Varela 1980; Varela *et al.* 1974). An autopoietic system—the minimal living organization—is one that continuously produces the components that specify it, while at the same time realizing it (the system) as a concrete unity in space and time, which makes the network of production of components possible. More precisely defined: An autopoietic system is organized (defined as unity) as a network of processes of production (synthesis and destruction) of components such that these components:

(i) continuously regenerate and realize the network that produces them, and

(ii) constitute the system as a distinguishable unity in the domain in which they exist.

Thus, autopoiesis attempts to capture the mechanism or process that generates the *identity* of the living, and thus to serve as a categorical distinction of living from non-living. This identity amounts to self-produced coherence: the autopoietic mechanism will maintain itself as a distinct unity as long as its basic concatenation of processes is kept intact in the face of perturbations, and will disappear when confronted with perturbations that go beyond a certain viable range which depends on the specific system considered. Obviously, all of the biochemical pathways and membrane formation in cells, can be immediately mapped onto this definition of autopoiesis.

A different exercise—which I do not pursue here at all—is to see how this basic autopoietic organization, present at the origin of terrestrial life (Fleischaker 1988), becomes progressively complexi-

fied though reproductive mechanisms, compartmentalization, sexual dimorphism, modes of nutrition, symbiosis, and so on, giving rise to the variety of pro- and eukaryotic life on Earth today (Margulis 1981; Fleischaker 1988). In particular, I take here the view that reproduction is *not* intrinsic to the minimal logic of the living. Reproduction must be considered as an added complexification superimposed on a more basic identity, that of an autopoietic unity, a complexification which is necessary due to the constraints of the early conditions on a turbulent planet. Reproduction is essential for the viability of the living, but only when there is an identity can a unit reproduce. In this sense, identity has logical and ontological priority over reproduction, although not historical precedence.

We do not pursue here these historical complexifications, neither do I pursue another equally pertinent empirical question: Can a molecular structure simpler than the already intricate bacterial cell, satisfy the criteria of autopoietic organization? This question can be answered by two complementary approaches: (1) simulation and (2) synthesis of minimal autopoetic systems. There are advances in both fronts. As to the first, there some new results in the burst of work in artificial life, partly extending our early simulations in tesselation automata of (Varela *et al.* 1974). The second front, takes the form of a new 'cell-centered' approach to the origin of life which seeks chemical embodiments of minimal autopoietic systems. In fact, the encapsulation of macromolecules by lipid vesicles has been actively investigated as a promising candidate for an early cell (Deamer & Barchfeld 1982; Lazcano 1986; Baeza *et al.* 1987; see Deamer 1986). Luisi & Varela (1989) make the case that a reverse micellar system can come close to the mark for being a minimal autopoietic system. In particular, they discuss the case of a reverse micellar system hosting in its aqueous core a reaction which leads to the production of a surfactant, which is a boundary for the reverse micellar reaction. The interest of this case is that much is known about these chemical systems making it possible to actually put into operation a minimal autopoietic system. But I must leave these fascinating issues to return to my chosen topic here.

1.2.2 Identity of the living and its world

Autopoiesis addresses the issue of organism as a minimal living system by characterizing its *basic mode of identity*. This is, properly speaking, to address the issue at an ontological level: the accent is on the manner in which a living system becomes a distinguishable entity, and not on its specific molecular composition and contingent historical configurations. For as long as it exists, the autopoietic organization remains invariant. In other words, one way to spotlight the specificity of autopoiesis is to think of it self-referentially as that organization which maintains the very organization itself as an invariant. The entire physico-chemical constitution is in constant flux; the pattern remains, and only through its invariance can the flux of realizing components be ascertained.

I have addressed here only the minimal organization that gives rise to such living autonomy. As I have said, my purpose is to highlight the basic biologic which serves as the foundation from which the diversity visible in current organisms can be considered: only when there is an identity can elaborations be seen as family variations of a common class of living unities. Every class of entities has an identity which is peculiar to them; the uniqueness of the living resides in the kind of organization it has.

Now, the history of biology is, of course, marred by the traditional opposition between the mechanist/reductionists on the one hand and holist/vitalists on the other, a heritage from the biological problem-space of the XIXth century. One of the specific contributions of the study of self-organizing mechanisms—of which autopoiesis is a specific instance—is that the traditional opposition between the component elements and the global properties disappears. In the simple example of the cellular automaton illustrated above, it is precisely the *reciprocal causality* between the local rules of interactions (i.e. the components' rules, which are akin to chemical interactions) and the global properties of the entity (its topological demarcation affecting diffusion and creating local conditions for reaction) which is in evidence. It appears to me that this reciprocal causality does much to evacuate the mechanist/vitalist opposition, and allows us to move into a more productive phase of identifying various *modes* of self-organization where the local and the global are braided together explicitly through this reciprocal causality. Autopoiesis is a prime example of such dialectics between the local component levels and the global whole, linked together in reciprocal

relation through the requirement of constitution of an entity that self-separates from its background. In this sense, autopoiesis as the characterization of the living does not fall into the traditional extremes of either vitalism or reductionism.

A second, complementary dimension of basic biologic that is central to focus our discussion is the nature of the *relationship* between autopoietic autonomous unities and their environment. It is exhypothesis evident that an autopoietic system depends on its physico-chemical mileu for its conservation as a separate entity, otherwise it would dissolve back into it. Whence the intriguing paradoxicality proper to an autonomous identity: the living system must distinguish itself from its environment, while *at the same time* maintaining its coupling; this linkage cannot be detached since it is against this very environment from which the organism arises comes forth. Now, in this dialogic coupling between the living unity and the physico-chemical environment, the balance is slightly weighted towards the living since it has the active role in this reciprocal coupling. In defining what it is as unity, in the very same movement it defines what remains exterior to it, that is to say, its surrounding environment. A closer examination also makes it evident that this exteriorization can only be understood, so to speak, from the "inside": the autopoietic unity *creates a perspective* from which the exterior is one, which cannot be confused with the physical surroundings as they appear to us as observers, the land of physical and chemical laws *simpliciter*, devoid of such perspectivism.

In our practice as biologists we switch between these two domains all the time. We use and manipulate physico-chemical principles and properties, while swiftly shifting to the use of *interpretation* and significance as seen *from* the point of view of the living system. Thus a bacteria swimming in a sucrose gradient is conveniently analyzed in terms of the local effects of sucrose on membrane permeability, medium viscosity, hydromechanics of flagellar beat, and so on. But on the other hand the sucrose gradient and flagellar beat are interesting to analyze only because the entire bacteria points to such items as relevant: their specific significance as components of feeding behavior is only possible by the presence and perspective of the bacteria as a totality. Remove the bacteria as a unit, and all correlations between gradients and hydrodynamic properties become environmental chemical laws, evident to us as observers but devoid of any special significance.

I have gone into this lengthy harangue because I believe that this truly dialectical relationship is a key point. In fact, it might appear as so obvious that we don't appreciate its deep ramifications. I mean the important distinction between the environment of the living system as it appears to an observer and without reference to the autonomous unity—which we shall call hereafter simply the *environment*—and the environment *for* the system which is defined in the same movement that gave rise to its identity and that only exists in that mutual definition—hereinafter the system's *world*.

The difference between environment and world is the surplus of *signification* which haunts the understanding of the living and of cognition, and which is at the root of how a self becomes one. In other words, this surplus is the mother of intentionality. It is quite difficult in practice to keep in view the dialectics of this mutual definition: neither rigid isolation, nor simple continuity with physical chemistry. In contrast, it is easy to conflate the unit's world with its environment since it is *so* obvious that we are studying this or that molecular interaction in the *context* of an autonomous cellular unit, and hence to miss completely the surplus *added* by the organism's perspective. There is no food significance in sucrose except when a bacteria swims upgradient and its metabolism uses the molecule in a way that allows its identity to continue. This surplus is obviously not indifferent to the regularities and texture (i.e. the "laws") that operate in the environment, that sucrose can create a gradient and traverse a cell membrane, and so on. On the contrary, the system's world is built on these regularities, which is what assures that it can maintain its coupling at all times.

What the autopoietic system does—due to its very mode of identity—is to constantly confront the encounters (perturbations, shocks, coupling) with its environment and treat them from a perspective which is not intrinsic to the encounters themselves. Surely rocks or crystal beads don't beckon sugars gradients out of all the infinite possibilities of physico-chemical interactions as particularly meaningful—for this to happen a perspective *from* an actively constituted identity is essential. It is tempting, at this point, to slide into some vaporous clouds about "meaning" reminiscent of the worst kind of vitalism of the past or informational jargon of the present. What I emphasize here is that what is meaningful for an organism is precisely given by its constitution as a distributed process, with an indissociable link between local processes where an interaction occurs (i.e. physico-chemical forces acting on the cell), and the coordinated entity which is the autopoietic unity, giving rise to the handling of its

environment without the need to resort to a central agent that turns the handle from the outside—like an *élan vital*—or a pre-existing order at a particular localization—like a genetic program waiting to be expressed.

I would like to rephrase this basic idea by turning it upside down as it were. The constant bringing forth of signification is what we may describe as a permanent lack in the living: it is constantly bringing forth a signification that is missing, not pre-given or pre-existent. Relevance must be provided *ex nihilo*: distinguish relevant from irrelevant molecular species, follow a gradient uphill and not downhill, increase the permeability to this ion and not to that one, and so on. There is an inevitable contretemps between an autonomous system and its environment: there is always something which the system must furnish from its perspective as a functioning whole. In fact, a molecular encounter acquires a significance in the context of the *entire* operating system and of many simultaneous interactions.

The source for this world-making is always the breakdowns in autopoiesis, be they minor, like changes in concentration of some metabolite, or major, like disruption of the boundary. Due to the nature of autopoiesis itself—illustrated in the membrane repair of the minimal simulated example above—every breakdown can be seen as the initiation of an action on what is missing on the part of the system so that identity might be maintained. I repeat: no teleology is implied in this "so that": that's what the self-referential logic of autopoiesis entails in the first place. The action taken will be visible as an attempt to modify its world—change from place of different nutrients, increase in the flow of a metabolite for metabolic synthesis, and so on.

In brief, this permanent, relentless action on what is lacking becomes, from the observer side, the ongoing *cognitive* activity of the system, which is the basis for the incommensurable difference between the environment within which the system is observed, and the world within which the system operates. This cognitive activity is paradoxical at its very root. On the one hand the action that brings forth a world is an attempt to reestablish a coupling with an environment which defies the internal coherence through encounters and perturbations. But such actions, at the same time, demarcate and separate the system from that environment, giving rise to a distinct world.

The reader may balk at my use of the term cognitive for cellular systems, and my cavalier sliding into intentionality. As I said above, one of my main points here is that we gain by seeing the *continuity* between this fundamental level of self and the other regional selves, including the neural and linguistic where we would not hesitate to use the word cognitive. I suppose others would prefer to introduce the word "information" instead. Well, there are reasons why I believe this even more problematic. Although it is clear that we describe an X that perturbs from the organism's exteriority, X is not information. In fact, for the organism only is a *that*, a *some*thing, a basic stuff to in-form from its own perspective. In physical terms there is stuff, but it is for nobody. Once there is body—even in this minimal form—it becomes in-formed for a self, in the reciprocal dialectics I have just explicated. Such in-formation is never a phantom signification or information bits, waiting to be harvested by a system. It is a presentation, an occasion for coupling, and it is in this *entre-deux* that signification arises (Varela 1979, 1988; Castoriadis 1987).

Thus the term cognitive has two constitutive dimensions: first its *coupling* dimension, that is, a link with its environment allowing for its continuity as individual entity; second—by stretching language, I admit—its *imaginary* dimension, that is, the surplus of significance a physical interaction acquires due to the perspective provided by the global action of the organism.

1.3 Perception-action and basic neuro-logic

1.3.1 Operational closure of the nervous system

In the previous Section, I have presented the fundamental interlock between identity and cognition as it appears for a minimal organism. In this Section I want to show how the more traditional level of cognitive properties, involving the brains of multicellular animals, is in some important sense the continuation of the very same basic process.

The shift from minimal cellularity to organism with nervous system is swift, and skips the complexity of the various manners in which multicellular organisms arise and evolve (Margulis & Schwartz 1988; Buss 1987; Bonner 1988). This is a transition in units of selection, and one that implicates the somatic balance of differentiated populations of cells in an adult organism, as well as crafty development pathways to establish a bodily structure. As Buss has stated recently: "The evolution of development is the generation of a 'somatic ecology' that mediates

potential conflicts between cell and the individual, while the organism is simultaneously interacting effectively with the extrasomatic environment" (Buss 1987).

For most vertebrates, this "somatic ecology" is bound together through the network of lymphocytes that constitute the core of the immune system. Again, a discussion of an immunological self is not my purpose here. I cannot resist the temptation, nevertheless, to point out, for completeness sake, that elsewhere I have presented *in extenso* a network approach to the immune system and its role in the establishment of a flexible cellular/molecular self during the ontogeny of mammals (see Varela *et al.* 1988; Varela & Coutinho 1991). In my view this identity is not, as traditionally stated, a demarcation of self as defense *against* the non-self of invading antigens. It is a self-referential, positive assertion of a coherent unity—a "somatic ecology"—mediated through free immunoglobulins and cellular markers in a dynamical exchange. Immune reactions against infections, although clearly important, are mediated by a "peripheral " immune system, a different sub-population of lymphocytes mobilized not through network but clonal expansion mechanisms, like a reflex reactivity acquired through evolution. But enough of this *excursus.* For my purposes here I will expeditiously assume the identity of a multicellular organism, distinctly different from an autopoietic minimal entity in its mode of identity, but similar in that it demarcates an autonomous entity from its environment.

Now, what's the specific place of the nervous system in the bodily operation of a multicellular? Whenever *motion* is an integral part of the lifestyle of a multicellular, there is a corresponding development of a nervous system linking effector (muscles, secretion) and sensory surfaces (sense organs, nerve endings). *The fundamental logic of the nervous system is that of coupling movements with a stream of sensory modulations in a circular fashion.* The net result are perception-action correlations arising from and modulated by an ensemble of intervening neurons, the *interneuron* network. Correspondingly, neurons are unique among the cells of a multicellular organism in their axonal and dendritic ramifications permitting multiple contacts and extending for large distances (relative to cellular soma sizes) providing the essential medium for this intra-organismic sensor-effector correlation.

Contrary to current habit, I wish to emphasize from the start the *situatedness* of this neuro-logic: the state of activity of sensors is brought about *most* typically by the organism's motions. To an important extent, behavior is the regulation of perception. This does not exclude, of course, independent perturbations from the environment. But what is typically described as a "stimulus" in the laboratory, a perturbation which is deliberately independent of the animal's ongoing activity, is less pertinent (outside the laboratory) for understanding the biology of cognition.

The perceptuo-motor coherencies we describe externally as behavior disguises the arising, within the interneuron net, of a large sub-set—an *ensemble* as is usually said—of transiently correlated neurons. These ensembles are both the source and the result from the activity of the sensory and effector surfaces. What changes is the amount of mediating interneurons, and the specific architecture of the respective nervous system, containing various cortical regions, layers and nuclei. In humans some 10^{11} interneurons interconnect some 10^6 motoneurons which relate to 10^7 sensory neurons distributed in receptor surfaces throughout the body. This is a ratio of $10 : 100,000 : 1$ of interneurons mediating the coupling of sensory and motor surfaces. The rise and decay of neuronal self-organization, say, in the modest *Aplysia* siphon withdrawal (Zecevic *et al.* 1989) is all the more valid in larger brains. Thus for instance a study in the cat (John *et al.* 1986) finds that 5–100 million neurons are active throughout the brain during a simple visuo-motor task of pressing a lever. Such neural assemblies arise in a patchwork of regional areas, evincing the enormous distributed parallelism proper to vertebrate brains.

The neuronal dynamics underlying a perceptuo-motor task is, then, a network affair, a highly cooperative, two-way system, and not a sequential stage-to-stage information abstraction. The dense interconnections among its sub-networks entails that every active neuron will operate as part of a large and distributed ensemble of the brain, including local and distant regions. For example, although neurons in the visual cortex do have distinct responses to specific "features" of the visual stimuli (position, direction, contrast, and so on), these responses occur only in an anesthetized animal with a highly simplified (internal and external) environment. When more normal sensory conditions are allowed, and the animal is studied awake and behaving, it has become increasingly clear that the stereotyped neuronal responses to "features" are highly labile and context sensitive. These have been shown, for example, for the effect of bodily tilt or auditory stimulation. Furthermore, the response characteristics of most neurons in the visual cortex depend directly on other neurons localized far from their receptive fields (see

e.g. Allman *et al.* 1985); even a change in posture, while preserving the same identical sensorial stimulation, alters the neuronal responses, demonstrating that even the supposedly downstream *motorium* is in resonance with the *sensorium* (Abeles 1984).

If I may continue to use vision as an example, I can take the previous discussion up one level of generalization, to note that in recent years research has become the study, not of centralized "reconstruction" of a visual scene for the benefit of an ulterior homunculus, but that of a *patchwork* of visual modalities, including at least form (shape, size, rigidity), surface properties (color, texture, specular reflectance, transparency), three-dimensional spatial relationships (relative positions, three-dimensional orientation in space, distance), and three-dimensional movement (trajectory, rotation). It has become evident that these different aspects of vision are emergent properties of concurrent sub-networks, which have a degree of independence and even anatomical separability, but cross-correlate and work together so that a visual percept *is* this coherency.

This kind of architecture is strongly reminiscent of a "society" of agents to use Minsky's (1987) metaphor. This multi-directional multiplicity is counterintuitive but typical of complex systems. They are counterintuitive because we are used to the traditional causal mode of input-processing-output directionality. Nothing in the foregoing description suggests that the brain operates as a digital computer, with stage-by-stage information processing; such popular descriptions for a system with this type simply goes against the grain. Instead, to the network and parallel architecture corresponds a different kind of operation: there is a "relaxation" time of back and forth signals until everybody is settled into a coherent activity. Thus the entire cooperative exercise takes a certain time to culminate, and this is evident in that, behaviorally, every animal exhibits a natural temporal parsing. In the human brain this flurry of cooperation typically takes about 200-500 msec, the "nowness" of a perceptuo-motor unity. Contrary to what it might seem at first glance either ethologically or in our own introspection, cognitive life is not a continual flow, but is punctuated by behavioral patterns which arise and subside in chunks of time. This insight of recent neuroscience—and cognitive science in general in fact—is fundamental for it relieves us from the tyranny of searching for a centralized, homuncular quality to a cognitive agent's normal behavior.

Let me backtrack a moment and reframe our discussion on cognitive self alongside that of a minimal molecular self. I am claiming that contemporary neurosciences—like cell biology for the case of the living organization—gives enough elements to conceive of the basic organization for a cognitive self in terms of the *operational* (not interactional!) *closure* of the nervous system (Maturana & Varela 1980; Varela 1979). I speak of "closure" to highlight the self-referential quality of the interneuron network and of the perceptuo-motor surfaces whose correlations it subserves. The qualification "operational" emphasizes that closure is used in its mathematical sense of recursivity, and *not* in the sense of closed*ness* or isolation from interaction, which would be, of course, nonsense. More specifically, the nervous system is organized by the operational closure of a network of reciprocally related modular sub-networks giving rise to ensembles of coherent activity such that:

(i) they continuously mediate invariant patterns of sensory-motor correlation of the sensory and effector surfaces;

(ii) give rise to a behavior for the total organism as a mobile unit in space.

The operational closure of the nervous system then brings forth a specific *mode* of coherence, which is embedded in the organism. This coherence is a *cognitive self*: a unit of perception/motion in space, sensory-motor invariances mediated through the interneuron network. The passage to cognition happens at the level of a behavioral entity, and not, as in the basic cellular self, as a spatially bounded entity. The key in this cognitive process is the nervous system through its neuro-logic. In other words the cognitive self is the manner in which the organism, through its own self-produced activity, becomes a distinct entity in space, but always coupled to its corresponding environment from which it remains nevertheless distinct. A distinct coherent self which, by the very same process of constituting itself, configures an external world of perception and action.

1.3.2 Cognitive self and perceptual world

The nature of the *identity* of the cognitive self just discussed is, like that of the basic cellular self, one of *emergence* through a distributed process. The emergent properties of an interneuron network are, however, quite different in their properties and likely to be much more rich in possibilities. What I wish to emphasize here is recent insights into the easiness with which lots of simple agents having simple properties may be brought together, even in a

haphazard way, to *give rise* to what appears to an observer a purposeful and integrated whole, *without* the need for a central supervision. We have already touched on this theme when discussing the nature of the autopoietic process and cellular automata modelling, and later when discussing the constant arising and subsiding of neuronal ensemble underlying behavior. This issue of emergent properties is crucial for my whole argument here, although I base my conclusions on contemporary studies from various biology-inspired complex systems (Farmer *et al.* 1986; Langton 1989a).

What is particularly important is that we can admit that (i) a system can have separate local components which (ii) there is no center or localized self, and yet the whole behaves as a unit and for the observer it is as if there was a coordinating agent "virtually" present at the center. This is what I meant when referring to a *selfless* self—we could also postulate a virtual self: a coherent global pattern that emerges through simple local components, appearing to have a central location where none is to be found, and yet essential as a level of interaction for the behavior of the whole unity.

The import of such current models, formalisms and case studies of complex systems (i.e. emergent properties through coordinated simple elements) is, in my eyes, quite profound for our understanding of cognitive properties. It introduces an explicit alternative to the dominant computationalist/cognitivist tradition in the study of cognitive properties for which the central idea is that of syntax independent of materiality which can support a semantics for an environment. This is also becoming more and more true for the researchers of artificial cognitive systems, as the current connectionist schools have made it clear by now. What we find in brains is a promiscuous tinkering of networks and sub-networks giving no evidence for a structured decomposition from top to bottom as is typical of a computer algorithm. Accordingly, one of the first messages from the study of artificial neural networks in modern connectionist terms is the absence of a principled distinction between software and hardware, or more, precisely between symbols and non-symbols. In fact, *all* we find in modern artificial neural network machines are relative activities between ensembles underlying the regularities we call their behavior or performance. We may see that some of these ensembles recur regularly enough to describe them as being program-like, but this is another matter. Although artificially built, such emerging ensembles cannot be called "computations" in the sense that their dynamics cannot be formally specifiable as the implementation of some high-level algorithm. Neural networks even in their fine detail are *not* like a machine language, since there is simply no transition between such elemental operational atoms with a semantics and the larger emergent level where behavior occurs. If there were, the classical computer wisdom would immediately apply: ignore the hardware since it adds nothing of significance to the actual computation (other than constraints of time and space). In contrast, in distributed, network models these "details" are precisely what makes a global effect possible, and why they mark a sharp break with tradition in AI. Naturally this reinforces the parallel conclusions that apply to natural neural networks in the brain, as we discussed before.

I have raised this point to caution the reader against the force of many years of dominance of computationalism, and the consequent tendency to identify the cognitive self with some computer program or high level computational description. This will not do. The cognitive self *is* its own implementation: its history and its action are of one piece. Now this demands that we clarify now the second aspect of the self to be addressed: its mode of relation with the environment.

1.3.3 Intentionality and neuro-logic

Ordinary life is necessarily one of *situated* agents, continually coming up with what to do faced with ongoing parallel activities in their various perceptuo-motor systems. This continual redefinition of what to do is not at all like a plan, stored in a repertoire of potential alternatives, but enormously dependent on contingency, improvisation, and more flexible than planning. Situatedness means that a cognitive entity has—by definition—a perspective. This means that it isn't related to its environment "objectively", that is independently of the system's location, heading, attitudes and history. Instead, it relates to it in relation to the perspective established by the constantly emerging properties of the agent itself and in terms of the role such running redefinition plays in the system's entire coherence.

Again, as we did for the minimal cellular self, we must sharply differentiate between environment and world. And again the mode of coupling is double. On the one hand, such body-in-space clearly happens through the interactions with the environment on which it depends. These interactions are of the nature of macrophysical encounters—sensory transduction, muscle force and performance, light and radiations, and so on—nothing surprising about them.

However this coupling is possible only if the encounters are embraced *from the perspective* of the system itself. This amounts, quite specifically, to elaborating a *surplus signification* relative to this perspective. Whatever is encountered must be valued one way or another—like, dislike, ignore—and acted on some way or another—attraction, rejection, neutrality. This basic assessment is inseparable from the way in which the coupling event encounters a functioning perceptuo-motor unit, and it gives rise to an *intention* (I am tempted to say "desire"), that unique quality of living cognition (Dennett 1987).

Phrased in other terms, the nature of the environment for a cognitive self acquires a curious status: it is that which *lends itself* (*es lehnt sich an...*) to a surplus of significance. Like jazz improvisation, environment provides the "excuse" for the neural "music" from the perspective of the cognitive system involved. At the same time, the organism cannot live without this constant coupling and the constantly emerging regularities; without the possibility of coupled activity the system would become a mere solipsistic ghost.

For instance, light and reflectance (among many other macrophysical parameters such as edges and textures, but let us simplify for the argument's sake), lend themselves to a wide variety of color spaces, depending on the nervous system involved in that encounter. During their respective evolutionary paths, teleost fishes, birds, mammals, and insects have brought forth various different color spaces not only with quite distinct behavioral significance, but with different dimensionalities so that it is not a matter of more or less resolution of colors (Thompson *et al.* 1992). Color is demonstrably not a property that is to be "recovered" from the environmental "information" in some unique way. Color is a dimension that shows up only in the phylogenetic dialogue between an environment and the history of an active autonomous self which partly defines what counts as an environment. Light and reflectances provide a mode of coupling, a perturbation which triggers, which gives an occasion for the enormous in-formative capacity of neural networks for constituting sensori-motor correlations and hence to put into action their capacity for imagining and presenting. It is only *after* all this has happened, after a mode of coupling becomes regular and repetitive, like colors in ours—and others—worlds, that we observers, for ease of language, say color corresponds to or represents an aspect of the world.

A dramatic recent example of this surplus significance and the dazzling performance of the brain as the generator of neural "narratives" is provided by the technology of the so-called "virtual realities". Visual perception and motions thus give rise to regularities which are proper to this new manner of perceptuo-motor coupling. What is most significant for me here is the *veracity* of the world which rapidly springs forth: we *inhabit* a body within this new world after a short time of trying this new situation (i.e. 15 minutes or so), and the experience is of truly flying through walls or of delving into fractal universes. This is so in spite of the poor quality of the image, the low sensitivity of the sensors, and the limited amount of interlinking between sensory and image surfaces through a program that runs in a personal computer. Through its closure, the nervous system is such a gifted synthesizer of regularities that any basic material suffices as an environment to bring forth a compelling world.

This very same strategy of the situatedness of an agent which is progressively endowed with richer internal self-organizing modules is becoming a productive research program even for the very pragmatically oriented field of artificial intelligence. To quote R. Brooks, one of the main exponents of this tendency at some length:

> I ... argue for a different approach to creating Artificial Intelligence:
>
> - We must incrementally build up the capabilities of intelligent systems at each step of the way and thus automatically ensure that the pieces and their interfaces are valid.
> - At each step we should build complete intelligent systems that we let loose in the real world with real sensing and real action. Anything less provides a candidate with which we can delude ourselves.
>
> We have been following this approach and have built a series of autonomous mobile robots. We have reached an unexpected conclusion (**C**) and have a rather radical hypothesis (**H**).
>
> **C** : When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. It turns out to be better to use the world as its own model.
>
> **H** : Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems.

> Representation has been the central issue in Artificial Intelligence work over the last 15 years only because it has provided an interface between otherwise isolated modules and conference papers.
>
> Brooks (1987, p. 1)

When the synthesis of intelligent behavior is approached in such an incremental manner, with strict adherence to the sensory-motor viability of an agent, the notion that the world is a source of information to be represented simply disappears. The autonomy of the cognitive self comes fully in focus. Thus in Brooks's proposal his minimal creatures join together various activities through a rule of cohabitation between them. This is homologous to an evolutionary pathway through which modular subnetworks intertwined with each other in the brain. The expected result are more truly intelligent autonomous sense-giving devices, rather than brittle informational processors which depend on a pregiven environment or an optimal plan.

It is interesting to note that in this paper Brooks also traces the origin of what he describes as the "deception of AI" to the tendency in AI (and in the rest of cognitive science as well) to abstraction, i.e., for factoring out situated perception and motor skills. As I have argued here (and as Brooks argues for his own reasons), such abstraction misses the essence of cognitive intelligence, *which resides only in its embodiment.* It is as if one could separate cognitive problems in two parts: that which can be solved through abstraction and that which cannot be. The second is typically perception-action and motor skills of agents in unspecified environments. When approached from this self-situated perspective there is no place where perception could deliver a representation of the world in the traditional sense. The world shows up through the enactment of the perceptuo-motor regularities. "Just as there is no central representation there is no central system. Each activity layer connects perception to action directly. It is only the observer of the Creature who imputes a central representation or central control. The creature itself has none: it is a collection of competing behaviors. Out of the local chaos of their interactions there emerges, in the eye of the observer, a coherent pattern of behavior" (Brooks 1986, p. 11).

To conclude, the two main points that I have been trying to bring into full view in this Section devoted to the cognitive self are as follows. First, I have tried to spell out the nature of its *identity* as a body in motion-and-space through the operational closure of the interneuron network. This activity is observable as multiple sub-networks, acting in parallel and interwoven in complex *bricolages,* giving rise again and again to coherent patterns which manifest themselves as behaviors. Secondly, I have tried to clarify how this emergent, parallel and distributed dynamics is inseparable from the *constitution of a world,* which is none other than the surplus of meaning and intentions carried by situated behavior. If the links to the physical environment are inevitable, the uniqueness of the cognitive self is this constant genesis of meaning. Or, again to invert the description, the uniqueness of the cognitive self is this constitutive *lack* of signification which must be supplied faced with the permanent perturbations and breakdowns of the ongoing perceptuo-motor life. Cognition is action about what is *missing,* filling the fault from the perspective of a cognitive self.

This view amounts to a biology of intentionality. In fact, it answers without ambiguity two key problems: the symbol (Harnad 1991) and the syntax grounding problems (Searle 1990). The first one refers to the mystery of the origin of signification of natural symbols, since in the classical cognitivist option there is an intrinsic need for an arbitrary semantic assignment. The answer provided by this approach is that the signification arises in the emergence of a viewpoint proper to the autonomous constitution of the organism at all its level, starting with its basic autopoiesis. The syntax grounding problem claims that all syntactic operations in a symbol system are observer-dependent. Our answer is precisely that the constitution of an autonomous unit provides the means for regularities to appear which are the bases of composionality. This can manifest at the cellular level as with the celebrated genetic code for protein sysnthesis, or at the brain level with compositional properties of neural ensembles. There is nothing mysterious in the emergence of such composable regularities. Thus contrary to most philosophical debate today (be this Searle, Harnad, or Dennett) we do not need to have an arbitrary observer-dependent assiginment of either significance or compositionality. The key is in the identity properties generated by the self-constitution of the organism.

1.4 Organism's double dialectics

Organism, then, is a key center for cognitive science, and it cannot be broached as a single process. We are forced to discover "regions" that interweave in

complex manners, and, in the case of humans, that extend beyond the strict confines of the body into the socio-linguistic register.

Further, what I have argued is that behind this meshwork of the various selves we carry around, is that all of these selves share a common and fundamental logic while differing in their specificity. This is a case of what Wittgenstein would have called "family resemblances": rather than any characteristic being common to all instances, we deal with a *cluster* of overlapping characteristics. We may also speak of this cluster of common characteristics as a shared *dialectic*, since we are dealing here with double-sided process, where co-definition is at the core of the matter. In fact, I submit that the organismic dialectic of self is a two-tiered affair: We have on the one hand the dialectics of identity of self; on the other hand the dialectics through which this identity, once established, brings forth a world from an environment. Identity and knowledge stand in relation to each other as two sides of a single process: that forms the core of the dialectics of all selves.

First, a *dialectics of identity* establishes an autonomous agent, a for-itself (*pour soi*). This identity is established through a bootstrapping of two terms:

(i) a *dynamical* term which refers to an assembly of components in network interactions and which are capable of emergent properties: metabolic nets, neural assemblies, clonal antibody networks, linguistic recursivity;

(ii) a *global* term which refers to emerging properties, a totality which conditions (downwardly) the network components: cellular membranes, sensory-motor body in space, self/non-self discrimination, personal 'I'.

These two terms are truly in a relation of co-definition. On the one hand the global level cannot exist without the network level since it comes forth through it. On the other hand the dynamical level cannot not exist and operate as such without it being contained and lodged into an encompassing unity which makes it possible.

Second, a *dialectics of knowledge* establishes a world of cognitive significance *for* this identity. This can only arise from the perspective provided by this identity, which adds a surplus of significance to the interactions of the environment proper to the constituting parts.

The key point, then, is that the organism brings forth and specifies its own domain of problems and actions to be "solved"; this cognitive domain does not exist "out there" in an environment that acts as a landing pad for an organism that somehow drops or is parachuted into the world. Instead, living beings and their worlds of meaning stand in relation to each other through *mutual specification* or *co-determination*. Thus what we describe as significant environmental regularities are not external features that have been internalized, as the dominant representationalist tradition in cognitive science—and adaptationism in evolutionary biology—assumes. Environmental regularities are the result of a conjoint history, a congruence which unfolds from a long history of co-determination. In Lewontin's (1983) words, the organism is both the subject and the object of evolution.

This second tier of the organism's dialectics, then, is also established through the bootstrapping of two terms:

(i) a *significance* term which refers to the necessary emergence of a surplus meaning proper to the perspective of the constituted self: cellular semantics, behavioral perception and action, self/non-self as somatic assertion, personal identity,

(ii) a *coupling* term which refers to the necessary and permanent embeddedness and dependency of the self on its environment, since only through such coupling can its world be brought forth: physico-chemical laws for the cellular world, macroscopic physical properties for cognitive behavior, molecular interaction for immune self, socio-linguistic exchanges for our subjective selves.

Double dialectics: the nature of an identity and the nature of a relation to a world. Double paradoxicality: Self-production by dependent containment; autonomy of knowledge through environmental coupling. Both dialectics give rise to the shifting nature of organism, ineluctably forming itself and in-forming where it is, and equally ineluctably implicated in the background from whence it springs forth. Organisms, those fascinating meshworks of selfless selves, no more nor less than open-ended, multi-level circular existences, always driven by the lack of significance they engender by asserting their presence.

Acknowledgments

The financial support of CNRS, Fondation de France (Chaire Scientifique) and the Prince Trust Fund is gratefully acknowledged.