# LA NUOVA CRITICA

RIVISTA DI SCIENZE DELL'UOMO E DI FILOSOFIA DELLE SCIENZE

# 2015

# NUOVA SERIE

# I-II

# QUADERNO 65-66

# AUTOPOIESIS, REFLEXIVITY AND THE GENESIS OF MIND

# IN MEMORY OF FRANCISCO VARELA

FRANCISCO J. VARELA, Self-organization: beyond appearances and into the mechanism
LOUIS H. KAUFFMAN, Mathematical themes of Francisco Varela
MICHAEL BITBOL AND PIER LUIGI LUISI, Autopoiesis with or without cognition: defining life at its edge
ARTURO CARSETTI, The emergence of the enactive mind in ambient meaning
HUMBERTO R. MATURANA, L'illusione della percezione, ovvero la chiusura operativa del sistema nervoso 103

Direttore responsabile: Arturo Carsetti

Pubblicazione semestrale iscritta al n. 17 del Registro della stampa di Roma (14 gennaio 1987)

Tipografia Union Printing, SS Cassia Nord km 87, Viterbo

Nel 1982 La Nuova Critica offri al lettore italiano una prima presentazione sufficientemente organica delle idee e delle direzioni di analisi che avevano sino a quel momento animato la ricerca portata avanti da H. Maturana e da F. Varela. Il Quaderno (La Nuova Critica, N. 64, 1982) dal titolo: "Autopoiesi e teoria dei sistemi viventi" conteneva scritti di A. Carsetti, H. Maturana, R. Uribe e F. Varela. La presentazione fu concordata direttamente con Francisco Varela da colui che scrive che curò sul piano scientifico insieme all'amico scomparso l'impostazione del quaderno stesso.

A distanza di tanti anni e proprio nel momento in cui in varie parti nel mondo si è proceduto o si procede a rivisitare in senso critico la eredità di F. Varela con pubblicazioni di grande respiro che vedono il concorso anche di molti e validi studiosi delle ultime generazioni, La Nuova Critica ha ritenuto opportuno procedere anch'essa ad una rivisitazione mirata di una delle eredità scientifiche tra le più importanti del secolo scorso, una rivisitazione che vuole, innanzitutto, essere un omaggio al lavoro di una vita dedita in modo esemplare a portare avanti un lavoro di ricerca a carattere interdisciplinare tra i più impervi tra quelli messi in cantiere nella seconda metà del Novecento. Molte sono le radici da cui Varela si diparte e molti sono i risultati raggiunti. È pressoché impossibile immaginare di racchiudere nelle poche pagine di un Quaderno anche una semplice e rudimentale catalogazione ragionata degli argomenti sottoposti a disamina da Varela lungo il corso della sua vita. Matematico, cibernetico, biologo, filosofo, studioso della cognizione e dei sistemi neurali, epistemologo, neurofenomenologo ecc. egli anche per tramite della sua straordinaria interdisciplinarietà è venuto aprendo in continuazione nuovi orizzonti della ricerca, orizzonti che si stanno sempre più rivelando, al momento attuale, ai nostri occhi di contemporanei tra i più fruttiferi.

Data la vastità di quella che è l'eredità scientifica di F. Varela si è deciso di limitare la rivisitazione critica qui proposta ad alcuni, po-

chi temi di fondo: il tema dell'Autopoiesi, il tema della Riflessività ed il tema concernente la genesi della mente e si è fatto ricorso in vista di approfondire in chiave critica tali temi a lavori mirati ad opera dei Proff. M. Bitbol, A. Carsetti, L. Kauffman e P. L. Luisi. Il Quaderno si apre con la riproposizione dell'articolo di Varela scelto a suo tempo dallo stesso Varela per la presentazione pubblicata nel 1982 ad opera de La Nuova Critica e da lui tradotto personalmente dal francese in inglese, articolo originariamente destinato ad essere pubblicato in francese nel 1983 con il titolo : "<u>Varela</u>, F., L'auto-organisation: De l'apparence au mécanisme" nel volume a cura di: P.Dumouchel e J.P.Dupuy, L'Auto-Organisation: De la physique au politique (*Eds. du Seuil, Paris, pp.147-165*).

Si tratta di una riproposizione importante per varie ragioni anche perché tale traduzione in inglese non essendo facilmente reperibile neppure sul web, risultava richiesta da più parti a livello dei paesi anglofoni. In particolare, il lavoro di Varela pubblicato in anteprima da La Nuova Critica si rivela come una adeguata sintesi della poliedricità della ricerca portata avanti da Varela negli anni in cui egli si accingeva ad operare il suo trasferimento definitivo in Europa. Serrato, ad esempio, appare in questo lavoro il confronto scientifico con le idee ed i metodi portati avanti da Prigogine. Va aggiunto, tuttavia, come a seguito della pubblicazione del quaderno Varela, al suo ritorno in Europa e nel mentre che si trovava come visiting professor presso l'Università di Roma, provvide a consegnare a colui che scrive un lavoro scientifico di grande importanza da lui scritto in Cile in collaborazione con un matematico nel periodo immediatamente precedente il ritorno stesso (Cfr. Andrade, J. and F. Varela (1984), Selfreference and fixed points, Acta Applic.Matem . 2:1-19).

Fu, appunto, in quella occasione che, a seguito di un invito a lui rivolto (ed esteso ad altri amici tra cui Vittorio Somenzi) da Valerio Tonini per una riunione conviviale, si giunse ad accennare da parte di colui che scrive, ma anche da parte di altri, ad un possibile trasferimento in Italia del grande studioso cileno. Si trattò, in realtà, di una idea soltanto accarezzata a livello della mente da parte di alcuni. A ragion veduta, il trasferimento in Europa che rappresentava per Francisco un desiderio molito sentito non poteva che avvenire in terra di Francia, grazie anche alla lungimiranza di Dupuy ed alla maggiore articolazione nel settore da parte delle strutture d'Oltralpe.

A seguito del suo trasferimento in Francia e del lavoro svolto presso il CREA, Varela venne a sviluppare una mole di lavoro che lascia ancora oggi stupefatti. Particolarmente importanti appaiono le ricerche da lui portate avanti nei campi concernenti l'enactive Realism e l'enactive mind, ricerche che a partire dalla prima impostazione offerta dal grande studioso cileno sono venute, oggi, a far parte dei territori d'elezione di una molteplicità di ricercatori per quel che riguarda l'epistemologia e le scienze cognitive. Di ciò costituiscono una riprova, tra l'altro, sia i recenti Quaderni della Rivista Constructivist Foundations sia la collezione dei lavori di Varela da poco pubblicata in Francia a cura di M. Bitbol et al. (Le cercle createur: ecrits (1976-2001) / Francisco Varela) sia il volume ormai in procinto di essere stampato di A. Carsetti dal titolo: Metabiology. Non-standard Models, General Semantics and Natural Evolution. Una ulteriore riprova è rappresentata dal fatto che le suddette intuizioni di Varela concernenti il Realismo enactive, (intuizioni che riprendono a partire dall'articolo sopra citato alcuni geniali apporti teorici da parte di Prigogine così come alcune più antiche osservazioni di H. von Foerster) vedono, oggi, a giudizio di colui che scrive, una loro, sia pur del tutto limitata e puramente speculativa, convergenza con alcune delle tesi proprie del participatory Realism così come divisato da alcuni brillanti fisici teorici che lavorano a livello della Quantum Mechanics ma nel solco di alcune primitive intuizioni di B. de Finetti (a partire dal suo lavoro fondamentale: B. De Finetti (1937), La prévision: ses lois logiques, ses sources subjectives, Ann. Inst. H. Poincaré, 7: 1-68).

A chiusura del quaderno e come una sorta di backstage, si è, altresì, deciso di riproporre il testo scelto da Maturana ed editato da Varela in vista di presentare al pubblico italiano nel suddetto Quaderno de La Nuova Critica del 1982 i primi fondamenti della dottrina della Autopoiesi.

La Nuova Critica desidera esprimere il proprio ringraziamento più vivo ai Proff. M. Bitbol, A. Carsetti, L.H. Kauffman e P.L. Luisi per aver accettato l'invito loro rivolto a partecipare alla "costruzione" del presente Quaderno.

AC

6

FRANCISCO J. VARELA

# SELF-ORGANIZATION: BEYOND APPEARANCES AND INTO THE MECHANISM

«It takes two to invent anything. The one makes up combinations, the other one chooses, recognizes what he wishes and what is important to him in the mass of the things the former has imparted to him. What we call genius is much less the work of the first one than the readiness of the second one to grasp the value of what has been laid before him and to chose it».

PAUL VALERY<sup>1</sup>

In a meeting devoted to self-organization, such as this one, it might seem a bit odd to present a proposal stating that self-organization is not a phenomenon, but rather an epiphenomenon of something deeper about a *specific class of systems*. This something deeper is *the mechanism which defines them as a class of systems*. Further, I will argue, it is only by developing our understanding and our analytical tools about this class of organization and its consequences, that we will make and headaway in this area. Let me unfold these ideas in a step-by-step fashion.

#### Units

1. I want to start from Valery's words, for they are admirably direct to the point I want to reach in the end. He meant them for imagination and poetic creativity. I do not think he would mind if we bring them in a discussion about self-organization and novelty in nature.

1.1. The first deep notion in these words is the necessity of the number two. Indeed, in order to talk about self-organization, one is already assuming the differentiation between a unit (or system if you wish) and its ambient (or background if you wish).

1.1.1. This is immediately seen by trying to apply the predicate "self-organize" to items such as: the air around me or, the first two miles of ocean. Since we do not have a clear operational criteria for distinguishing a unit in either one of those cases, it is odd to predicate about them that they have a given organization, and even more so, that they could self-organize.

1.2. The moral of this boundary condition on our mode of cognizing is that we can separate a unit from its background whenever it makes sense to distinguish (at least) two *independent series of events*. That is to say, two domains of events which we can treat as if they had a measure of independence. For, what sense would it make to distinguish something from not-something, if there was only a network of interconnected events where we could agree on no separable cracks?

## Coupling

2. From Valery's starting intuition we have isolated the notion that whatever self-organization is, it entails the distinction of a unity and its background, and that these two stand in relation to each other as two series of events with a measure of independence. Now, what could this "measure" possible mean?

In fact, this is not an easy question to answer, because we are so used to think in terms of what I would like to call now «pointwise connected» independent series of events (or states of affairs). Let me explain. 2.1. Let us take the extremes first of all. On the one hand, we have the extreme of not being able to differentiate between unit and background. Thus, as we already concluded, questions about organization are senseless. On the other hand, we have a complete un-coupling between unit and its ambient, a complete adiabatic causality (as it were). This case might be interesting for Gedanken experiments, but cannot have any bearings in our understanding of natural systems or man-made artifacts.

2.2. It is thus evident that the formulation of interest is where we speak of a degree or partial independence between the events proper to the unit (or we would not distinguish it) and the events proper to the ambient (or we could not tell there is one). Unit and ambient are coupled at *some* points. In other words: there is a coupling surface at which mutual influences criss-cross, but such coupling surface is not the whole unit, but only one (or a few) dimensions of it. I shall abbreviate by saying that there is *pointwise coupling*.

2.2.1. Let us examine just a bit more closely this pointwise coupling. First of all, this is not all unfamiliar, for we use it everyday. In fact, we have a most explicit paradigm of pointwise coupling from systems theory: an input which modulates an ongoing dynamics of states.

2.2.1.1. More explicitely one defines a space of allowable inputs, I, a space of states, S, and dynamics f which tells us what the next state will be, given an input and a current state:

$$\begin{split} f: T \times I^m \times S^n & \to S^n & 1 \leq n, \, m \leq {}^\infty \\ (i,s) & f \\ t & \mapsto {}^st + \Delta t & \Delta t, \, t \epsilon T, \, i \epsilon I^m, \, s \epsilon S^n \end{split}$$

2.2.1.2. In this case, the pointwise coupling is made explicit by choosing a given domain of inputs which are permissible through specific modes of actions explicit in the transition law f. At the same time, independence (a degree of it) is made explicit by the transition function being dependent on internal states. This standard system-

theoretic paradigm<sup>2</sup> captures well the requirements of clear differentiation between unit and environment and their pointwise coupling. I do not need to elaborate how useful and applicable such a framework has been for the engineer and the physicist. I shall call this an *inputtype* coupling of the two series of events.

2.2.2. Input-type coupling is *one* mode of framing pointwise coupling, but certainly not the *only* one. For, what are we to do for those situations whose fundamental quality is that of autonomy? By autonomy I mean, at this point, an exclusive notion: those units for which I cannot find (or seems unsatisfactory to find) an input-type coupling descriptions.

2.2.2.1. Put in a more positive sense, autonomy means that the system is heavily tilted towards an internal determination or selfassertion. I have argued extensively elsewhere<sup>3</sup> that autonomy is a fundamental quality needed to understand natural systems such as the cell and the multicellular organisms, as well as organismic subsystems such as the immune and the nervous system.

2.2.2.2. Does this mean that we cannot treat autonomous systems rigorously? Surely not. It is a matter of where we put the emphasis. For, in an input-type coupling, we assume that the *points of contact* between the two independent series of events (the unit and not the unit) can serve as the *main guideline* for the future dynamics of that system. For an autonomous system the converse is true: the *internal transformations* are the main guiding thread to understand the system dynamics, and the points of coupling are needed only in so far we need contingencies to understand particular courses of transformation.

2.2.2.1. Please do note that I am putting the emphasis on what we need in order to come up with a satisfactory explanation. To see something as autonomous or not depends on the questions we want to answer. But when it comes to biology, we have the whole phenomenology of living systems to convince us that we get lost unless we concentrate on keeping track of the dynamics of internal transformation, and leave the points of coupling as perturbing agents rather than inputs. 2.2.2.2.2. What is the difference between inputs and perturbations? Basically that an input specifies the only way a given transformation of state by occur. A perturbation leaves unspecified the agent, but only takes into account its effect of that agent in the unit's structure.

In other words: an input enters in the definition of that unit. A perturbation can couple to it, but is not definitory. Thus a given perturbation can be brought about in indefinitely many different ways. A given input can only occur in its specified manner (that is why it was defined in the first place).

2.2.2.2.3. More explicitely, let us define a state space S, and an internal dynamics

$$f: T \times S^n \to S^n$$
  
 $l < n \le \infty, s \in S^n, t \in T$   
 $f^{s}t \mapsto f(t + \Delta t)$ 

The system will operate continuously until a perturbation hits it (which could be self-inflicted). The effect of such perturbation is to move the current state *and* the dynamics to a new configuration.

$$\begin{split} (f + \delta f) &: T \times S^n \to S^n \qquad t \epsilon T; \ s, \ s + \Delta S \epsilon S^n, \\ (s + \Delta s)_t & \stackrel{(f + \delta f)}{\longmapsto} {}^s (t + \Delta t) \qquad \qquad f, f + \delta f \epsilon \ [S^n \to S^n] \end{split}$$

where  $s + \Delta f$  and  $f + \delta f$  express, respectively, a perturbation in the state space to a new state, and a change of the previous dynamics to a new one in the appropriate space of endomorphisms of  $S^n$ , written here as  $[S \rightarrow S^n]$ .

Notice that in this case, the allowable perturbations are defined by the system's structure, since they can be *any* thing which leads to a change of state and/or dynamics. A study of these kinds of system naturally leads to consider the kinds of recurrent behavior capable of withstanding a history of perturbations. In contrast to the previous input-type systems, I call these autonomous systems. A specific example is offered below (3.2.2.1.).

2.2.2.2.4. Autonomy, therefore, is an attitude of characterizing a unit by its internal coherence in order to account for its identity and its history of coupling. (This central role of the internal coherence can be made more explicit though the notion of *operational closure*, as I have described before<sup>4</sup>, but we do not need to state those details here). To abbreviate I shall call this *closure-type coupling* of two series of events.

2.2.2.3. Thus, in considering the pointwise coupling we have at least two fundamental optics with which to look at our units: input-type and closure-type. Is this list exhaustive? I do not know, and it is certainly a very interesting question that merits more detailed dissection that I can afford here<sup>5</sup>. But herein might lie a key to better understand those units which are stubborn and resist either description.

## The origin of diversity

3. The next link in my chain of reasoning is a rather natural one. Once we concentrate on closure-type couplings, the immediate question is to find ways to inquire about the *kinds* of internal coherence that a class of units *can* exhibit.

3.1. To the extent that our emphasis is on internal relations and their determinations, and not through definitions by input (or outputs), it is their coherence which are the most proper specifications for such units. This is why I have used the name *eigenbehaviour*, or self-determined behaviours, for such internal coherences.<sup>6</sup>

3.2. The richness and thus the complexity of a system is directly expressible as the intricacy of the landscape of its possible eigenbehaviours. The dynamics in this landscape of eigenbehaviour will depend on the workings of the systems, and its peculiar pathways on its history of perturbations.

12

3.2.1. We may be more specific on this point, as it is necessary. We do this by assuming that we may associate to an eigenbehaviour an empirical observable or measurement. Let f, g be observations adequate to a system with states in a space S. When observed under an f-optics the system's states will become an equivalence classes under f, that is,  $S/R_f$ . Correspondingly under g, other equivalence classes will be induced in S,  $S/R_g$ :



Assume furthermore that we may compare our f-optics with our g-optics under some suitable metric for S. Then define a class  $(s)_f$  as stable under g if and only if  $\pi_g \pi^{-1}{}_f(U)$  is an  $\epsilon$ -neighborhood of  $(s)_g$  whenever U is a  $\delta$ -neighborhood of (S)f.

Now, the complement of the set of stable classes of  $S/R_f$  with respect to  $S/R_g$  will be called the *bifurcation* set of f with respect to g.<sup>7</sup>

3.2.2. This general notion of bifurcation is a way of saying what it is for a system to look "very" different under two perspectives. That is, a way to have a sense for how intricate an internal coherence is, is to measure how many different optics yield irreducible aspects of the system. It is a very adequate precision of the notion of *complexity* (or richness or diversity if you wish) which hunts any consideration of self-organization.

3.2.2.1. Let me illustrate. Consider an autonomous unit (in the sense of 2.2.2.2.3.) specified in a fairly simple (in some sense simplest) way: iteration of the same action. Concretely, let the state space be the closed interval [-1,1], and dynamics be a function  $f \in FC$  [S  $\rightarrow$  S], where F is a sub-space of the continuous endormophisms of S, [S  $\rightarrow$  S] spanned by a parameter  $\mu$  in

$$f(x) = 1 - \mu x^2 \quad x \in [-1,1], \ \mu \in [0,2]$$

Our system can now characterized by an iteration of this dynamics,

$$\begin{array}{ccc} f & f: IN \times S \rightarrow S \quad [-1,1] = S, & n \epsilon IN = T \\ & & f \\ & & x_n \quad \xrightarrow{f} \quad x_{(n+1)} = f(x_n) = 1 - \mu x_n^2 \end{array}$$

In this apparently simple case, it is clear that the coherent states or eigenbehaviours will be the fixed points of some period m, i.e. those values  $x^*$  such that

$$x^* = f^m(x^*), \qquad f = f, \text{ fof } = f^2$$

However, we may ask whether this system will look alike when we observe  $x^*$ , if the system is perturbed in the dimension of its parameter  $\mu$ . (Notice that one could rephrase this by saying: how approximate can our measurement of  $\mu$  be for the unit to have the same eigenbehaviour in its state  $x^*$ ?). That is, let us consider perturbations.

$$\delta \xrightarrow{f + \delta f} (f + \delta f) : IN \times S \rightarrow S$$

$$\overset{f + \delta}{x_n} \xrightarrow{f + \delta} x_{(n+1)} = 1 - (\mu + \delta) x_n^2$$

where  $\delta$  is some small number,  $\delta \ll \mu$ 

The answer, as it has been recently investigated<sup>8</sup> is surprisingly complex, for the bifurcation set (in the sense of 3.2.1.) is extremely rich. In fact, depending on the range of  $\mu$ , it might look like a branching tree of x<sup>\*</sup> of periods which grow like 2<sup>n</sup>.



Or more amazingly, in an appropriate range, a arbitrarily small change in  $\mu$ , will lead to an arbitrarily large change in the behaviour in S. (Some of these behaviours can reach infinite periods, and it is thus indistinguishable from noise from the point of view of the observation under that meter).

3.2.2.2. This example is very interesting for two reasons. First, because it says that closure-type units of some of the simplest dynamics can very easily get very complex, in the strict sense of richness of its bifurcation set.<sup>9</sup> Secondly, it should not escape the reader that these examples only studied so recently, could have been studied long ago (in fact, one needs only a pocket calculator to get to some of the main results). They were not, I submit, for the good reasons that theory always follows interest, and we have had a long neglect of an interest for closure-type coupling in favour of input-type couplings so prevalent in physics and engineering.

#### A mechanism for self-organization

4. I have now come to the core of my argument, which may be compactly stated as follows:

Every self-organizing behaviour is generated by the diversity in the internal coherence of an operationally closed system.

4.1. Properly speaking there are no self-organizing systems, only self-organizing behaviour. A self-organizing behaviour leads us to the study of the mechanisms capable of generating diversity. (One approximation to this diversity is available through the notion of a bi-furcation set).

Self-organization is a description of a behaviour which is useful and heuristic to point to a phenomenon. But it is bound to stay *only* as a behavioral description unless one searches for the mechanisms of its origins. This is the direction in which my proposal moves.

4.2. Note that an input-type view of a system does not, per se, exclude a study of self-organization of that system. But since that perspective will not, in general, be interested in the internal coherence but in the ways in which inputs drive the system, it will simply not look for them or such description obscure those possibilities.

4.2.1. In contrast, if we do put an emphasis in the self-organizing quality of a system we autonomatically assume a closure-type stance towards it, and we are therefore looking at a different unit, although later, we may englobe both perspectives in a broader view.

4.2.2. This explains to me why it is that there has been a tendency to confuse self-organization with operational closure (or, conversely, to distinguish between self-organizing system and control system). For example, it has been said that an autopoietic system<sup>10</sup> is another way of saying that cells self-organizing.<sup>11</sup> This misses the point for it would be like saying that, for example, intentional goal-seeking behaviour is saying the same thing as an optimization algorithm. The two descriptions – a behaviour and a mechanism – are related for the observer to the extent that he may superimpose both views on a larger perspective. But this does not change the fact that one describes behaviour which could be produced by potentially many different mechanisms, and the other specifies a mechanism which uniquely specifies a behaviour.

4.3. Another point I wish to make is that my proposal to move from behaviour into mechanisms in this area, is consistent with its history. As we know, the idea of self-organization grew out of the study of those system which could produce a changing adaptive be-

haviour under conditions of constant perturbations or just random noise<sup>12</sup>.

4.3.1. Note, in the first place, that the very idea of sharply separating the system and its working from the environment by considering it as a source of perturbations or noise, immediately points to a change of view as I have describe before: from an input-type stance to a closure type stance. This was definitively a step in the right direction, although it was never, in my opinion, completed. We need to complement it with a more radical view of a study of eigenbehaviours and their diversity.

4.3.2. I believe that my proposal also clarifies the role of noise in this whole area of thinking. There is the sense that noise is something like the main course in the menu of self-organizating system. They eat it, digest it, and produce order. From our vantage point things look rather differently.

4.3.2.1. Noise enters into the picture only to the extent that we observers, take a stance whereby the interesting or relevant aspects to describe are those that have to do with the internal coherence of the system. Therefore, the structure of the ambient is left unspecified, and thus it appears (or it is assumed as if) random. But this is not to say that noise has "something" which feeds the system. This would be precisely to fall in the input-type stance again. The apparent mystery dissipates as soon as we understand how, in a given system, its closure can generate a diversity of eigenbehaviours. Which eigenbehaviours were triggered by which perturbations may or may not be of our interest, but the key is not in the nature of the perturbations, but in the fact that whatever they are they can bring about changes in the eigenbehaviour.

## Learning and Evolution

5. The two self-organizing situations par excellence have been taken to be learning and evolution. That is, a situation typical of an individual ontogeny, and a situation proper to a supra-individual unit, that of a population. Let me briefly sketch how a closure-stance to both these situation changes quite radically our approach to both of them. With these illustrations, I mean to follow my arguments in the sense that the distinctions I have drawn before *do* matter and are not idle. These examples are my experiments, as it were, to substantiate my hypothesis.

5.1. First, let me characterize the two viewpoints about the brain which amount in take either an input or a closure-type stance.

5.1.1. The input-type stance amounts to treat the nervous system as being defined fundamentally through *inputs*. The inputs in this case are taken to be *features* or qualities of the environment which should be taken up as the raw material in be processed inside. In brief: the nervous system works with an information content of the ambient, and works on this information by making an operational *representation* of it.

5.1.2. The closure-type stance amounts to treat the nervous system as being defined fundamentally through the internal coherences which are attained by its relative interconnectivity.<sup>13</sup> More precisely stated, by the eigenbehaviour which are produced through the mutual (neuroanatomical) *mappings* between its various internal surfaces together with the local actions proper to each one of them. Sensory and motor actions are, from this pointview, only one link in an ongoing operational closure, albeit an important one for the observer of behaviour.

5.1.3. The reader will agree with me that these two characterization amount to see rather different systems, and to look for rather different things in the experimental designs. By and large the representational view is more predominant today, (and it has been over the last fifty years<sup>14</sup>). Yet, there is the closure-type alternative view which seems, to me, both more interesting and more simple. As I cannot develop in full this point of view, I will have to make do with a few remarks to motivate it, lest the reader thinks it is totally crazy to say that the nervous system works without inputs and outputs. The representational view needs no motivation, since all texts book are built around it.

5.1.3.1.Consider the realm of visual experience. It seems natural, at a first glance (so to say), to agree that the world appears as full of

textures which by simple actions, like moving my head, are neatly taken in by my perception. This first-order experiential view seems to suggest naturally an input-type stance and to look for features of the environment which are picked up by various stages in my retina, then on up to the visual cortex.

5.1.3.2. However, this first impression of our experience is misleading. A closer look reveals that indeed what I see has more to do with the way I am put together as a mechanism that with an out-there.<sup>15</sup> This more subtle view of experience is in fact easier to correlate with the structure of the nervous system.

5.1.3.2.1. Thus, for example, whatever happens at the level of my retina as a sensory surface has only one effect: modulation of neural activity in the several neuronal aggregates to which it connects on retinotopic basis. One of them is the lateral geniculate nucleus (LGN) of the thalamus, which the text calls "a relay station to the cortex". However, as one fiber from the retina comes in at a certain position in the LGN, at least five other comes in at the same location.<sup>16</sup> As a result, whatever the effect of the retinal activity is, it cannot but act as only a modulator of something which is ongoing at the inside of the nervous system. In this case what is relevant is the state of relative activity between the various interconnects of the LGN with other brain layers, and through them, with the whole brain. A diagram may help to visualize the situation:



#### Francisco J. Varela

5.1.3.2.2. This kind of system is more naturaly (or more easily) described as one whose internal coherence, attained through the reciprocal layer-to-layer interconnections, is *modulated* via a coupling surface, such as the retina. But the key in the system's operation is the attainment and diversity of its eigenbehaviour, not the nature of the modulation. For example, we tend to think of colour as an attribute of objetcs. However, at closer examination color is virtually independent (except in very restricted situation) from the illumination that reaches the eye. Colour is defined for us, experientially, through a mechanism to which we have no direct experiential access. Such mechanism essentially consists of an operation of relative comparison between levels of activity, and the invariants obtained through this sort of mechanism *do* correlate well with our experience of colours.<sup>17</sup>

5.1.4. If we shift our stance towards the nervous system from an input to a closure-type, then what we look for as the mechanism of learning and memory undergo a corresponding change.

5.1.4.1. One of the universal properties of all known nervous tissues is that their structure at the local and cellular level is plastic. The conclusion immediately follows: given a system with operational closure, whose structure changes slightly under a history of perturbations, to an observer it will look it has learned and that it keeps a record of what has happened. For the system however, no such record is necessary, for a given change can happen at many sites simultaneously and has no direct relationship with the perturbing agent. It is only for the observer that, if the perturbation is recurrent, this event will look like a recognition, computers work by storage and retrieval; nervous systems, however, do not.

5.1.4.2. This is not to say that we can say now with accurate details exactly how do the landscape of eigenbehaviours change with the small scale changes the cellular synaptic level. Work in this direction is just beginning. We need a better understanding of how to characterize an eigenbehaviour,<sup>18</sup> and how new eigenbehaviour can arise under perturbations.<sup>19</sup> This is certainly a major area to work in.

5.2. Let me now turn in a corresponding characterization of the two viewpoints about evolution which amount to take either an input or a closure-type stance.

20

5.2.1. The input-type *stance* amounts to treat the environment as the main guideline to understand the dynamics of the modifications by descent and its genetics, the result being a better *adapted* lineage of organisms. In other words, natural selection is an example of an optimization algorithm; the survival of the fittest.

5.2.2. The closure-type stance amount to treat the internal coherences of an animal population as the guiding tread in understand phyletic transformation. The result is that there is the generation of diversity but no sense of optimization of adaptation. In fact, adaptation is an *invariant*, just like identity is an invariant as long as the organisms does not disintegrate.<sup>20</sup> Natural selection is here only a description of the major boundary conditions, the range which such phyletic diversity can proceed.

5.2.3. The reader will, again, agree with me that these two stances amount to rather different views of evolution, and they imply rather different things to look for and in designing experiments. Interestingly enough, the input-type stance or adaptationist paradigm has also been in dominance over the past fifty years (at least in the anglo-saxon science).<sup>21</sup> Yet there is today, and there has been for quite sometime (in fact, since Darwin himself), the alternative closure-type stance. Again, the textbook view needs no motivation, for we are all used to it. Lets the reader thinks, however, that to abandon an adaptationist viewpoint is as suicidal as abandoning a representational view of the brain, let me say a few work of motivation for this view.

5.2.3.1. Consider the bodyplan of any organism. The adaptationist tendency is to see this body as a collection of traits whose design can be explained though an optimal fit with the corresponding aspect of the environment. At best such traits are seen as having some cross correiations where a cost-benefit analysis can apply. However, this result is still a collection of traits. It is thus that it comes to be that we have heard the stories on how did the fish acquired its fins and the kangaroo its big feet: it is a superior design for what it is meant for.

5.2.3.2. However, this first impression is, again, misleading. For it leaves out the most striking fact about both development and or-

ganismic physiology, namely, that the unit does not work as a set of traits but as a coherent whole. The position of a limb in an embryo cannot be understood as some natural provision for what is to come next but as the result of the interdependence and mutual definition of what goes on in an *embryo* at each point. To change one point, is to change radically its eigenbehaviour, and just one trait.

5.3.2.3. Thus an important continental school<sup>22</sup> holds that the constraints in a body plan which arise from this operational closure of an organism "restrict possible paths and modes of change so strongly that the contraints themselves become the most interesting aspect of evolution"<sup>23</sup>.

For example A. Seilacher<sup>24</sup> notes how the beautiful and intricate patterns of the shells of certain molluscs are most easily explained as diversity which arises from the diversity of materials to satisfy an invariant mode or architectural development. Correspondingly, he argues, these different forms do not correspond to some sought-for optimal adaptation.

5.3.2.4. This is not to say, again, that we have all the details of how such mechanism work. In fact, at this stage evolutionary theory is in the middle of a profound revision. However, independent of the kinds of explicit mechanisms that are refined to express these ideas, it is evident that the change of our stance from one of optimum design and adaptation, to one of evolution ad tinkering,<sup>25</sup> is a very profound one.

## Conclusion

Let me present to you the bare skeleton of the logic I have followed in this article. It has 5 steps:

- 1. Self-organization is a behaviour which is proper to autonomous units.
- 2. Autonomous units can be characterized if we change from an input-type stance to a closure-type stance.

- 3. Closure-type requires the understanding of the internal coherence (eigenbehaviours) such units have.
- 4. When a unit exhibits enough structural plasticity its coherences will be diverse and complex.
- 5. Such diversity in self-determined internal coherences, when observed as behaviour under the appropriate contingencies of interaction, will appear as novelty, unpredictability, self-assertion, in brief, as an autonomous self-organizing unit.

This proposed mechanism constitues, in fact, an explanation of the phenomenon of self-organization because it is capable of *generating* it. I discussed how this perspective leads to a change in our current understanding of at least two important situations: learning and evolution. Our quest was into how the proposed mechanism actually operates in both cases.

I wish to draw this presentation to an end with three commentaries.

#### First

It is no curious coincidence that for the last few decades, in two major areas of biology – neurobiology and evolution – an input-type description was chosen in favour of a (perfectly accessible) closure-type stance. I believe this touches on a very deep prevalent contemporary sensibility of our contemporary science. The best way I know of formulating it is by saying that it favours a *representational epistemology*, that is, the predominant view of knowledge as a world-picture.<sup>26</sup> Specially in the context of anglosaxon scientific philosophical tradition, this assumption feels like home.

In contrast, I firmly believe that there is a major change, or trend of a change in our contemporary sensibilities and scientific epistemology in the sense that we are becoming more and more interested in a different epistemology, which is not concerned with the world-as-picture, but with the *laying-down* of a world, where a unit and its world co-arise by mutual specification.

#### Francisco J. Varela

It could be said that the notion of self-organization serves as a clear symptom that differentiates between input-type machines (whether we call them Turing, state-transition, or simple functionalist theories), and biological autonomy and human understanding. Because self-organizing behaviour depends on a history of coupling, and it is based on a mechanism which is explicitly distributive and hermeneutic: it is interpretative, precisely, in the sense of laying down of being.

It is in the sharing of this sensibility, and only in this sense, that I see that there is a commonality between very disparate lines of research in various fields, and which has been lumped under the term self-organization, englobing from physics<sup>27</sup> to language and communication.<sup>28</sup> I remain somewhat skeptical of grand synthesis of all of them, for it fails to pinpoint the relevant commonality, which is epistemological, and to search instead for a unit of methodology, which is at best misleading.

#### Second

As for my second closing comment. Note that in both my outlines for how neurobiology and evolutionary thinking change in changing our stance, it is evident that such a change of stance is not a contradictory rejection. Again, here I see the possibilities of a rather new styles of thinking relative to what has been typical in natural sciences, with few glorious exceptions.

Consider for example evolutionary thinking and the adaptationist programme. It is true that change a stance gives a different view of life. However, the selectionist stance need not be negated, and in fact it is a much simple explanation for some changes in the body plan of an animal lineage. A typical example would be the tremendous environmental changes during the triasic and their impact in the flora and fauna that survived. On the same vein, for neurobiology, courtship between two birds seems more easily explained as a typical case of release of a very stable mechanism by an input-type action, the sexual partner.

24

Thus, at least for these two stances discussed here, it is possible to superimpose the kinds of system they generate, so that we may have a higher-level complementary view. I find this possibility very interesting, but also there is the danger that lies in not keeping a clear track of the kind of stances we are taking in each case, which lead us into believing that we are addressing the very same system.

Thus, my plea is not for exclusivity of one view but for a measured pluralism. This is to say: a pluralism which keep clear track of what are the actions we did that lead into seeing one or the other kind of system, and how are we to superimpose them into a more englobing perspective.

I do not know whether we might not need other stances which are as distinct and cleanly definable as the ones I have outlined here. Maybe we do need them, and they could be appropriate for those systems, notably the social situations, which seem yet to resist attempts to dig our conceptual teeth deep into them.<sup>29</sup>

However, an explicit methodology for multidimensional complementary seem to me one of the interesting directions in which our many apparently disparate lines of research seem to be leading us. The need for such clear-headed multidimensionality is one of the trade marks that biological and social system impose on our science.

#### Third

My third and last closing commentary concerns the question of novelty. In fact, it is impossibile to think about self-organization without at the same time addressing this fascinating issue. I shall do so now from the frame I have outlined before.

One question seems to me fully answered, namely: Where does novelty come from? I have taken the position that novelty is always the result of the specific kind of mechanism outlined here; in principle, we know where does all novelty come from. This, however, raises another underlying question: Once I have a mechanism which accounts for the origin of a behaviour, is it still novel? Can I call the results of a knowable mechanism novel? So, apparently we have answered one question only to fall into another, equally interesting one: Is there such as thing as true novelty? My answer to this question is no, yes. *No*, there is no novelty if the situation we are dealing with looks new because we did not have at the time enough insight into the dynamics of the unit in question. Such novelty vanishes whenever we improve our study and come up with an appropriate dynamics, which is the origin of such behaviour. No, there is no novelty if the situation we are dealing with looks new because we do not have a record of the history of perturbation which make sense of the transformation from one kind of behaviour to the supposedly new one. Such novelty vanishes whenever we improve our record (or our patience) and come up with the appropriate history of contingencies. In a word: novelty is not truly such if is bound to ignorance or partiality in perspective.

But, *yes*, there is novelty if the very act of asking for the system's behaviour, and thus entering in some interaction with it, changes its dynamics. This is novelty which will not vanish, because we have entered into a new dynamics, us and the observed behaviour, which in itself may exhibit self-organizing behaviour. But this new, higher-order dynamics is not separable from the act of asking the question.

We are all familiar with such interesting entanglement from quantum mechanisms. But this is only the tip of the iceberg compared with the situation we find in trying to find mechanisms for our *own* understanding. Since the very activity of theory making is of the same order of what it wants to explain, it will inevitably be incomplete. We cannot step outside our cognitive domain specified by our biological and social beings. All we have is the current state, and there will be, of necessity, an irreducibie realm of interactions; perturbations always appear shifty, ungrounded, and unpredictable.

In my answer to the question about novelty then, novelty takes on a different sense. Not that which is not accessible to mechanisms or description. But that which falls in the inevitable *cognitive blind spots* of our current understanding and experiencing.

Novelty is interesting, not because it is inaccessible to our descriptions, but because the logic of the situation is such that we fall within the goedelian vortices of entangled hierarchies. It is in the nature of such closure-type mechanisms that true novelty arises, of the kind that has to do with human experience of freedom and creativity. And Paul Valery, is apparently, of the same mind.

#### Notes

- 1 P. VALERY as quoted by J. HADAMARD, *The Psychology of Inventing in the Mathematical Field*, Princenton, U. Press, 1949, p. 30 ff.
- 2 For a clear statement of this view of systems theory see L. PADULO and M. AR-BIB, *System Theory*, Saunders, Philadelphia, 1974.
- 3 F. VARELA, *Principles of Biological Autonomy*, North-Holland, New York and Oxford, 1979.
- 4 In F. VARELA (1979), *op. cit.*, Chapter 7. The term used in this book is organizational closure. Subsequently I have found that the same idea is better expressed by using the term *operational* closure.
- 5 Sec GOGUEN and VARELA, "Systems and distinctions, duality and complementarity", *Int. J. Gen. Systems 5:* 1 (1978), *op cit.*, and F. VARELA, "Not one, not two", *CoEvolution Quart.*, Fail Issue (1976), for more on the notion of complementarity and its formalization through adjointness in categories.
- 6 The explicit definition of eigenbehaviour can be found in F. VARELA (1979), *op. cit.*, Chapter 13.
- 7. This formulation of a bifurcation set is due to ROBERT ROSEN, *Fundamentals of Measurement and Representation in Natural Systems*, North-Holland, New York and Oxford, 1978, pp. 37-45.
- 8 See the clear account of R. MAY (1976), "Simple mathematical models with very complicated dynamics", *Nature*, *261*: 459-467.
- 9 For an account of how interesting and universal this complexity is, see O. LANFORD (1980), "Smooth transformations of intervals", *Seminaire Bourbaki* n. 563, Oct. 1980.
- 10 The definition of an autopoietic system is to be found in H. MATURANA and F. VARELA, *"Autopoiesis and Cognition"*, D. Reidel, Boston, 1980.
- 11 As an example see M. ZELENY (1977). "Self-organization of living systems: a formal model of autopoiesis". *Int. J. Gen. Systems* 4: 13.

#### Francisco J. Varela

- 12 I am thinking for example, of the volume edited by M. YOVITS, G. YOVITS, G. JACOBI and G. GOLDSTEIN, *Self-organizing Systems*, Spartan Books, Washington, D.C., 1962.
- 13 For a full statement this view of the nervous system, see MATURANA and VARELA (1980), op. cit.
- 14 See for example the introductory article by D. HUBEL in the Sept. 1979 issue of *Scientific American*, entirely devoted to the brain. Another good example is the introductory chapter of the popular textbook by S. KUFFLER and J. NICHOLS, *From Neuron to Brain*, Sinauer Associates, Boston, Mass., 1977.
- 15 To say more about this point would take us too far afield. But the work of Merleau-Ponty is probably the best source.
- 16 The sources of this information are many, but I have been specially influenced by W. SINGER (1977), *Physiol. Rev.* 57: 386-420.
- 17 For details on this issue, but from a rather different perspective, see E. LAND and J. McCANN (1971), "Lightness and retinex theory". *J. Opt. Soc. Am.* 61: 1:11.
- 18 One form such eigenbehaviours can take is, for example, what Freeman calls a «macrostate». See W. FREEMAN, *Mass Action in the Nervous System*, Academic Press, New York, 1975.
- 19 One form such history of interactions can take is, for example, what is called "associative memory". See T. KOHONEN, *Associative Memory: A system-theoretic Approach*, Springer-Verlag, Berlin, 1977, and the more neurobiology-oriented account of J. ANDERSON (1979), "Parallel computation with simple networks", *Cognition and Brain Theory 3*: 45-53.
- 20 MATURANA and F. VARELA (1981). "Evolution: Natural drift and the principle of conservation of adaptation » (forthcoming).
- 21 This section is substantially based on S. GOULD and R. LEWONTIN (1979), "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist paradigm", *Proc. Roy. Soc. Ser. B. 205*: 581-598.
- 22 See for example the book by the french zoologist P. P. GRASSÉ, *Evolution of Living Organisms*, Academic Press, New York, 1978.
- 23 S. GOULD and R. LEWONTIN, op. cit., p. 594.
- 24 A. SEILACHER (1972), "Divaricate patterns in pelecypod shells", *Lethaia 5*: 325-343.
- 25 This apt way of phrasing the issue is due to F. JACOB (1977), "Evolution as tinkering", *Science 196:* 1161.
- 26 I owe this insight to work of the continental hermeneutical school. See in particular M. HEIDEGGER's "Die Zeit des Weitbildes" in *Holzwege*, V. KLOSTER-MAN, Frankfurt, 1952.

- 27 I am, of course, thinking of the pioneering work in physical-chemistry of I. PREGOGINE and his group (e.g. G. NICHOLS and I. PRIGOGINE, *Self-Organization in Non-equilibrium Systems*, J. Wiley, New York, 1977), and M. EIGEN and his collaborators (e.g. M. EIGEN and P. SCHUSTER, *The Hypercycle*, Springer-Verlag, Berlin, 1979).
- 28 See for example E. JANTSCH, *The Self-Organizing Universe*, Pergamon Press, New York, 1980.
- 29 C. CASTORIADIS, *La Institution Imaginaire de la Societé*, Seuil, 1975. For further discussion on these issues and, in generally, in the study of social systems and autonomy, I have been specially influenced by J.P. DUPUY & P. DU-MOUCHEL *L'Enfer des Choses*, Sevil, Paris, 1980.

LOUIS H. KAUFFMAN\*

# MATHEMATICAL THEMES OF FRANCISCO VARELA

#### I. Introduction

This is an article about the possibilities in the mathematical work of Francisco Varela. The themes of this article will be related to my personal history of work with Varela since that work is directly involved in the mathematics he used and the possibilities inherent in that mathematics. In this introduction I shall summarize in a nutshell the mathematical ideas that Varela and I worked on together and I shall give some facts and opinions about their development. Many of these ideas were recorded by Varela in his seminal book "Foundations of Biological Autonomy" (Varela (1979).

Francisco was one of the early people to recognize the significance of the work "Laws of Form" (G. Spencer-Brown (1969) hereafter denoted LOF. He saw that LOF, based on the fundamental idea of distinction, articulated a crucial concept that is foundational for biology and the biology of language. By starting with a distinction we understand that it is through a distinction that an organism makes, in the eye of an observer, that the organism exhibits structural stability,

<sup>\*</sup> Mathematics, University of Illinois at Chicago, Chicago, Illinois 60607-7045, <Kauffman@uic.edu>

autonomy and becomes an exemplar of the living. This notion of distinction is crucial to our understanding of the nature of an organism and the nature of life itself. The distinction is a joint creation of the organism in its environment and the observer. Together they give life to the organism. The distinction does not appear without the observer, and the distinction that is the organism does not appear without the actions of the organism, producing itself from itself through components taken from and given back to the environment. A crucial model of this epistemology is given by Maturana, Uribe and Varela in their paper "Autopoesis - the organization of living systems - its characterization and a model" (Maturana, Uribe and Varela (1974)).

The notion that an organism must have autonomy of structure, and yet that structure is intimately related to the interaction with the environment and with an observer is the theme of Varela's book "Principles of Biological Autonomy". Varela took a very daring intellectual step in this book by creating a new extension of LOF that included at its base a symbol for autonomy. This is the symbol of the reentering mark:



Let the letter J denote this reentering mark. The reader can take this symbol at the allegorical level as an Ouroboros, a World-Snake, the Snake that Eats its Tail. The mythical indication of the self-devouring, self-creating Universe.

Just alongside the mythological level, we have the notion of the form that reenters its own indicational space. The form sits inside itself, just as the WorldSnake sits inside itself. We can indicate this reentrance by an equation

$$J = \langle J \rangle$$
  
 $J =$ 

where the brackets indicate the outer boundary of the reentering mark, and we see that J sits inside itself. Using the original reentrant mark we write



The device of Barthélemy Aneau, wood block print from *Picta poesis*, 1552 (Image from Glasgow University)

In this way we indicate that J can appear as an infinite nest, and so appear identically within itself. I will say more about this theme later in the essay, but we should remark here that the infinite nesting solution J = <<<<...>>> requires discussion of the nature of infinity and it certainly is a solution that goes outside the finite forms that a person might search for in looking for a solution to the equation J =<J>. The point about the "autonomous form" J is that it can be seen to cross the boundary between object and process. It is at once both an object and a process, just as a biological organism is both an object and a process in the eye of its beholder. Varela took the step of identifying such an autonomous form at the basis for his "calculus for selfreference" and he placed it in alignment with the autonomous nature of a living organism.

#### II. Meeting and Work with Varela

I discovered the book "Laws of Form" [G. Spencer-Brown (1969)] in 1974, some time after the book had been published, and two years after I had completed a PhD in mathematics from Princeton University. I was teaching at the University of Illinois at Chicago Circle (as it was called at that time). I encountered a special experience in the foundations of thought and mathematics, almost as soon as I picked up that book by G. Spencer-Brown. Spencer-Brown's book was a turning point in my intellectual life. Laws of Form is a lucid exposition of the foundations of mathematics. It embodies a movement from creativity, to creation, to symbol, to system and language and thought and self. Expressing that creation took away the apparent ground of my previous conception. There was no longer any distinction between the certainty or uncertainty of mathematics and the certainty or uncertainty of present experience. There was no longer any distinction between geometry/topology and logic. There was no longer any possibility that logic could be the foundation of mathematics, or that mathematics could have any foundation other than itself. There are many roads to this place. For me, Laws of Form came along at the right time. Later, I enjoyed reading accounts of similar experiences with Laws of Form by non-mathematicians such as Alan Watts and John Lilly.

A year after my encounter with Laws of Form a seminar arose at the Circle Campus (now called the University of Illinois at Chicago). This seminar, devoted to Laws of Form, met every Wednesday at the apartment of Kelvin Rodolfo, a professor of geology at the Circle. The seminar met cordially for the whole evening in a comfortable living room with food and drink and ample time for discussion. We were an eclectic group: David Solzman from Geography, Rachel MacKenzie, Gerry Swatez from Sociology, Joy Swatez, Kathy Crittenden from Sociology, Mike Lieber from Anthropology, Paul Uscinski a young computer scientist, Brayton Gray from Mathematics, myself and others. We read the book, argued over it, free associated to it, performed it and generally wandered in the space opened from the possibility of a single distinction. This seminar had a life of more than three years and it deeply influenced the lives of all its members.

Paul and I became fascinated by the recursive and circuitous world of Chapter 11 in Laws of Form. We invented for ourselves an interpretation of the workings of those circuits, and we found ourselves writing the reentering mark to express these ideas.



I discovered, around this time, an article in the Whole Earth Magazine about a young biologist, Francisco Varela, who had just written a paper [Varela (1975)] about Laws of Form and had worked out an algebra that included the reentering mark! We found Francisco's paper and added it to the discussion in the seminar. I resolved to get in touch with him.

A little research turned up Francisco's connection with Heinz von Foerster and the Biological Computer Laboratory at the University of Illinois at Urbana-Champaign, about 150 miles from Chicago. We had earlier in the seminar called up Heinz to tell him that we were studying Laws of Form. Heinz had written the brilliant review of Laws of Form that appears in the Whole Earth Catalog, where he characterizes it as "Spencer-Brown's transistorized 20th century version of Occam's razor". When we called Heinz and told him of our endeavor he laughed and we laughed over the telephone.

I was fascinated by the notion of imaginary boolean values and the idea that the reentering mark and its relatives, the complex numbers could be regarded as such values. *The idea is that there are "logical values" beyond true and false, and that these values can be used to prove theorems in domains that ordinary logic cannot reach.* At that time I was fascinated by the reentering mark, and I wanted to think about it, in and out of the temporal domain.

The reentering mark has a value that is either marked or unmarked at any given time. But as soon as it is marked, the markedness acts upon itself and becomes unmarked. "It" disappears itself! However, as soon as the value is unmarked, then the unmarkedness "acts" to produce a mark!

You might well ask how unmarkedness can "act" to produce markedness. How can we get something from nothing? The answer in Laws of Form is subtle. It is an answer that destroys itself. The answer is that any given "thing" is identical with what it is not. Thus markedness is identical to unmarkedness. Light is identical to darkness. Everything is equivalent to nothing. Comprehension is identical to incomprehension. Any duality is identical to its confusion into union. There is no way to understand this "law of identity" in a rational frame of mind. An irrational frame of mind is (in this view) identical to a rational frame of mind. All is the working of the reentering mark. In Tibetan Buddhist logic there is existence, nonexistence and that which neither exists nor does not exist [T. Stcherbatsky (1968)]. Here is the realm of imaginary value.

The condition of reentry, carried into time, reveals an alternating series of states that are marked or unmarked. This primordial waveform can be seen as an alternation of plus and minus or as an alternation of minus and plus:

··· + - + - + - + - + - + - + - ... , ··· - + - + - + - + - + - + - + ...

or as

Marked, Unmarked, Marked, Unmarked,... Unmarked, Marked, Unmarked , Marked,...
I decided to examine these two total temporal states as representatives of the reentering mark, and I called them I and J respectively [L. H. Kauffman (1978) – DeMorgan Algebras]. These two imaginary values fill out a world of *possibility* perpendicular to the world of true and false.



I wrote a paper about I and J, showing how they could be used to prove a completeness theorem for a four valued logic based on True, False, I and J. I called this the "waveform arithmetic" associated with Laws of Form. In this theory the imaginary values I and J participate in the proof that their own algebra is incomplete. This is a use of the imaginary value in a process of reasoning that would be much more difficult (if not impossible) without it. Prior to that I had written a paper using Francisco's "Calculus for Self- Reference" to analyze the temporal behavior of self-referential circuits [L.H. Kauffman (1978) – Network Synthesis]. My papers were inspired by Varela's use of the reentering mark in his analysis of the completeness of the calculus for self-reference that he associated with that symbol.

I also started corresponding with Francisco, telling him all sorts of ideas and recreations related to self-reference. We agreed to meet, and I visited him in Boulder, Colorado in 1977. There we made a plan for a paper using the waveform arithmetic. This became the paper "Form Dynamics", eventually published in the Journal for Social and Biological Structures [L.H. Kauffman and F. Varela (1980)]. An earlier attempt to publish it in the International Journal of General Systems was met by the criticism that we had failed to acknowledge the entire(!) Spanish School of Polish Logic. I still have the letter from that referee. Later I learned to appreciate Spanish Polish logic (a group of logicians in Spain working on DeMorgan algebras and related matters in multiple-valued logic). Francisco based a chapter of his book "Principles of Biological Autonomy" on form dynamics. I remember being surprised to find some of my words and phrases in the pages of his book.

The point about Form Dynamics was to extend the notion of autonomy inherent in a timeless representation of the reentering mark to a larger context that includes temporality and the way that time can be implicit in a spatial or symbolic form. *Thus the reentering mark itself is beyond duality, but implicate within it are all sorts and forms of duality from the duality of space and time to the duality of temporal forms shifted in time from one another, to the duality of form and nothingness "itself*". I believe that both Francisco and I felt that in developing Form Dynamics we had reached a balance in relation to these dualities that was quite fruitful, creative and meditative. It was a wonderful aesthetic excursion into basic science.

This work relates at an abstract level with the notions of autonomy and autopoesis inherent in the earlier work of Maturana, Uribe and Varela [H. Maturana, R. Uribe and F. J. Varela (1974)]. There they gave a generalized definition of life (autopoesis) and showed how a self-distinguishing system could arise from a substrate of "chemical" interaction rules. I am sure that the relationship between the concept of the reentering mark and the details of this earlier model was instrumental in getting Francisco to think deeply about Laws of Form and to focus on the Calculus for Self-Reference. Later developments in fractal explorations and artificial life and autopoesis enrich the context of Form Dynamics.

At the time (around 1980) that Francisco and I discussed Form Dynamics we were concerned with providing a flexible framework within which one could have the "eigenforms" of Heinz von Foerster [Heinz von Foerster (1981)] and also the dynamical evolution of these forms as demanded by biology and by mathematics. It was clear to me that Francisco had a deep intuition about the role of these eigenforms in the organizational structure of biology. This is an intuition that comes forth in his books [Varela (1974, 1975, 1979, 1981, 1987, 2001)] and in his other work as well.

There is a more general theme that has been around since that time. It is the theme of "unfolding from a singularity" as in catastrophe theory. In the metaphor of this theme the role of the fixed point is like the role of the singularity. The fixed point is an organizing center, but it is imaginary in relation to the actual behaviour of the organism, just as the "I" of an individual is imaginary in relation to the social/biological context. The Buddhists say that the "I" is a "fill- in". The linguists point out that " I am the one who says "I"." The process that is living never goes to the fixed point, is never fully stable. The process of approximation that is the experiential and experienced I is a process lived in, and existing in the social/biological context. Mind becomes conversational domain and "mind" becomes the imaginary value generated in that domain. Heinz von Foerster [Heinz von Foerster (1981)] said "I am the observed relation between myself and observing myself."

The biological context is a domain where structural coupling and coordination give rise to mind and language. The fixed point is fundamental to what the organism is not. In the imaginary sense, the organism becomes what it is not.

Francisco invited me to participate in summer science seminars held at the Naropa Institute in Boulder Colorado in the early 1980's. We had a group of scientists and courses of lectures: linguistics (Alton

Becker, Kyoko Inoue), poetry (Haj Ross), poetry and linguistics (Haj Ross), geography (David Solzman), biology(Varela and Maturana), psychology(Eleanor Rosch), Laws of Form (Kauffman), constructive mathematics (Newcomb Greenleaf) and more. We talked and talked. I do not know how many of us also meditated, but the atmosphere of the Buddhist Institute provided a wonderful place for the gestation and exchange of ideas.

After those Naropa years we saw each other a few more times . Once we drove together from a cybernetics meeting (a Gordon conference) to a weekend retreat at the Buddhist center Karme Choling in Northern Vermont. I saw him again in Paris in 1989 and once at at conference in Brussels ("Einstein Meets Magritte") in the 1990's.

In that time I kept returning to Laws of Form and our shared ideas and continued these themes for many years up to the present. [Kauffman (1978, 1980, 1985, 1987, 1994, 1995, 2001, 2003, 2005, 2009, 2010, 2011, 2015, 2016)].

In those same years, from 1978 until the middle 1990's I had a long and complex correspondence with G. Spencer-Brown that culminated in my paper [Kauffman (2005)- Reformulating the map color theorem] about his approach to the four-color map theorem. These conversations also revolved around the nature of mathematics and the nature of the circuit structures in Chapter 11 of Laws of Form.

Much of my work on cybernetics and Laws of Form grows out of this interaction with Varela. The reader can consult more recent papers of mine to see how this work expands on eigenform and recursion and how a subject I call Iterants grows out of Form Dynamics. There is much more to be done in understanding these ideas that come from a cybernetic and biological stance.

# III. From Categories and Functors to Eigenforms

Another important theme in Varela's book is the use of categories and functors to study biological autonomy. A category is a very general mathematical notion that is based on the philosophy that things, objects arise and acquire meaning through their relations and relationships. Just so, in a category there are objects, but we are not necessarily told anything about their "internal structure". Indeed the objects may not have any internal structure, just as the idealized points in geometry have no internal structure. This is in direct contrast to set theory where the objects are sets, and except for the empty set, every set has internal structure in the sense of its members. In category theory there are *objects* (possibly void of internal structure) and *morphisms*, represented as directed arrows from one object to another object. Thus if A and B are objects in a category, then there may be a morphism f: A à B. We name the morphism by the letter f, as shown. This has the same form as the notation for a function from one set to another, but it can denote nothing more than a directional relation from A to B.

If I have a morphism  $f:A \rightarrow B$  and another morphism  $g:B \rightarrow C$ , then it is an axiom of category theory that there shall be a composite morphism  $fg:A \rightarrow C$ . You should think of fg as obtained by going from A to B by f and then going from B to C by g. (I will use the ordering fg here so that we go from left to right lexicographically to compose the morphisms.) We can depict this relationship as shown in Figure 1.



Figure 1. Composition of Morphisms

The next axiom of category theory concerns three morphisms.

f: A $\rightarrow$ B, g: B $\rightarrow$ C and h: C $\rightarrow$ D.

Now we have (fg)h and f(gh) by performing the composition of f and g first and then composing with h, or by performing the compositions of g and h first and then composing with f. See Figure 2.



Figure 2. Associations of Morphisms

The second axiom of category theory asserts that there is an equality of morphisms (fg)h = f(gh). We say that composition of morphisms is *associative*.

Finally, the third and last axiom of category theory asserts that every object A has an *identity morphism*  $1_A$ : A $\rightarrow$ A. This is a morphism from A to itself, such that any composition with it by another morphism, leaves the other composition unchanged. For example, if f: A $\rightarrow$ B, then  $1_A f=f$ .

Any directed graph G generates a category. We let the nodes of the graph be the objects and we make each edge of the graph into a morphism. We add an identity morphism at each node (object) and define each directed path in the graph to be a composite morphism. Paths are composed in the usual way by having two paths such that the start node of one is the end node of the other. The composite path is obtained by walking along the paths consecutively. Thus we obtain from a directed graph G, a category Pa(G), the *path category* of the graph. In fact there is a complementarity between graphs and categories in the following sense. We can consider the collection of all graphs as a category in its own right where the objects are individual graphs. Let us call this category of graphs **Graphs**. We can also consider a category of all (small) categories. Here the objects are categories and we will call this category of categories Categories. Then we have functors (maps of categories preserving their structure) Pa: **Graphs**→**Categories**, and **F: Categories**→**Graphs**. The first functor is our path method of converting a graph to a category. The second functor F *forgets* the category structure and just sees the objects as nodes and the arrows of the category as directed edges. This back and forth association between graphs and categories is an example of an adjoint pair of functors. The reader will, in fact, find a lucid description of adjoint functors in Varela's book "Principles of Biological Autonomy" in Chapter 10, page 97. Varela made a good beginning in this work on a categorical analysis of the meaning and possibility in the concept of complementarity at many levels. He used this formalism of categories to think about many specific aspects of biological organisms such as the way the immune system operates and at another level how mind, language and body interact. At another level complementarities occur between the very concepts of process and object and he managed to summarize these with a precision that was bolstered by the underlying category theory.

Category theory is a way to approach mathematics and its applications with attention to concept and meaning. Consider the simplest category. This category has one object O and one morphism  $1_O: O \rightarrow O$ that composes with itself to produce itself. In other words, this simplest category just has one entity and the barest sort of self-reference of this entity to itself, in the form of the morphism  $1_O$ .

What is the next simplest category with only one object? We can have another morphism f:  $O \rightarrow O$  and this time we will not put any restrictions on the compositions of f with itself. Then we have many generated morphisms: f, ff, fff, ffff, ... In fact we have an infinity of such morphisms, and they correspond to going around the arrow f that points from O to itself. This category embodies the concept that "self-reference is infinity in finite quise". The arrow f makes a reference from O to O, and the iteration of this reference gives an associated infinity. These compositions are the paths in the graph that is indicated by the one node O and the one directed edge f. Thus this simple category indicates the adjoint relationship between the path category and the simple graph of self-reference.



Figure 3. Self Reference in a Simple Category

We can go further than the usual category theory by considering the infinite composition J = fffffff... and now we have arrived at the first eigenform, for it is the case that fJ = J. J is unchanged under the application of f. The infinity generated by the self-reference has returned to as in the form of self-reference as an eigenform.

There is clearly much more to be done in using categorical ideas in understanding cybernetics and biological autonomy. This is an aspect of Varela's work (and his collaborations with Joseph Goguen) that should be pursued vigorously by mathematical oriented cyberneticists.

Nevertheless, in following the epistemology and the fundamental notions of cybernetics, one should, in my opinion, start not with categories, but with Laws of Form. In Laws of Form we have a simplest possible mathematical formalism, a symbol < > that represents a distinction between its outside and its inside. (Here I will use these brackets. In Spencer-Brown's book a partial box, the mark, is used to the same purpose.) Even without any further axioms for using the mark, one can proceed to find a multitude of iconic forms: < > , <<>>>, <<<>>>,...

Here I have only indicated the simple nesting that we have discussed earlier. By taking an infinite nest as in

we obtain the reentering mark with  $J = \langle J \rangle$ . Other self-referential forms can be constructed by similar recursions. For example we can have F so that  $F = \langle F \rangle$  F> and G so that  $G = \langle G G \rangle$ . We have almost at once an arithmetic of infinity and reentrance emanating from the idea of a distinction.

In simplicity we realize that the mark < > can be seen as the result of crossing from the unmarked state. Then the icon < > for the mark can be seen as an operator that transports (an observer!) from the unmarked inside to the marked outside. Seeing that mark as an operator, we can then see that < < > can be interpreted as a passage from a marked state (on the inside) and crossing from the marked state

44

yields the unmarked state. Thus < < > > represents the unmarked state and we can write

<< >> =

where I place nothing to the right of the equals sign since the icon for the unmarked state is no icon at all.

By the same token, one can understand that two adjacent marks

< > < >

stands just for the marked state since each mark can be seen as the name of the other mark. Thus we can write

< > < > = < >.

These two equations complete the construction of the mark as a logical particle and form the beginning of the mathematics and epistemology of Laws of Form.

To this day we continue to work on interweaving the fundamental simplicity of Laws of Form with the more complex levels of recursion, reentry, category theory, eigenform and other aspects of mathematical modeling.

It is in following the epistemological track of autonomy that Varela was led to consider mathematical domains that included distinctions (Laws of Form) recursions, category theory, lambda calculus and eigenforms. All of these mathematical domains have stood the test of time in relation to cybernetics and yet all of them are really at the beginning of a development that can bring forth the potential that cybernetics has a nexus where the observer and the observed are not one nor are they two. The cybernetic nexus where the observer plays hide and seek in the relationships that become an observed world is our subject/object matter in working with this field. Varela had the courage to go into the mathematics and begin a necessary exploration of formal language in relation to biology and the nature of the organism.

### **IV. Eigenform and Reflexivity**

I want to summarize very briefly here a way that I work with to explore how the mathematics of reflexivity is relevant to cybernetics. Francisco Varela considered this theme in his work as well, and the difference between the approach I shall outline and his, is more a matter of taste than anything else. The reader should compare with Chapter 13 of "Foundations of Biological Autonomy".

I say that a set D is a reflexive domain if every element of D is also a transformation of D to itself. Thus given an object or element a in D there is a corresponding function (morphism) a:  $D \rightarrow D$ . Here the role of a as a morphism is that it is a function from D to D. This means that if b is an object in D then ab will be a new object in D. We then further assume that if we define a mapping from D to D using these morphisms, then the new mapping corresponds to an element of D. For example, if I define fx = (ax)(bx) for any x, then we assume that there really is an element f in D with this very property. In the reflexive domain, every 'person' is also an 'agent', who transforms the whole space D to itself. There is a complementarity in a reflexive domain between its objects and its morphisms.

A key theorem about reflexive domains is this following result [Barendregt (1985)].

**Church-Curry Fixed Point Theorem.** In a reflexive domain D, every element F has a fixed point. That is, if F is in D then there is g in D such that Fg = g. In other words every F in D is the transformation for an eigenform that corresponds to F.

**Proof of the Theorem.** Define Gx = F(xx). Then there is a G in D with this property, by the assumption that D is a reflexive domain. Therefore GG = F(GG) follows, by letting G act on itself. But this means that we can take g = GG and then Fg = g. This completes the proof. //

The remarkable fact about this construction of eigenform is that it does not involve an excursion to infinity. Look at a special case. Define  $Gx = \langle xx \rangle$ .

Then  $GG = \langle GG \rangle$  and so if J = GG, we have  $J = \langle J \rangle$ .

I have produced the reentrant mark without the 'construction' of infinitely many marks as I did in the introduction, and as I did in the section on category theory where I went around the circle of self-reference infinitely many times.

With this concept of reflexive domain we can continue the discussion and extend it beyond biology to the many reflexive domains that appear to us once we adjust a cybernetic lens and keep the observer as an actor in the system that engulfs us.

### V. A Last Word

I would like to give Francisco Varela the last word(s) in this essay by quoting a passage from "Principles of Biological Autonomy". This paragraph is from 16.3 "Linguistic Domains and Conversations" p. 267.

"From another perspective, if we consider a conversation as a totality, there cannot be a distinction about what is contributed by whom. Linde and Goguen (1978) ... in their careful descriptions of the structure of discourse, ... found no evidence that the text, as a coherent entity, could be attributed to separate speakers, but it was an alloy of their participation, and exhibited rules and laws that are not reducible to the separate contributions. A similar basic methodological principle is behind Pask's approach to teaching machines, where a conversation is a coherent recursive aggregate.

... These ideas are precisely in line with the central theme of this book: that every autonomous structure will exhibit a cognitive domain and behave as a separate, distinct aggregate. Such autonomous units can be constituted by any process capable of engaging in organizational closure, whether molecular interactions, managerial manipulations, or conversational participation. I am saying, then, that whenever we engage in social interactions that we label as dialogue or conversation, these constitute autonomous aggregates, which exhibit all the properties of other autonomous units. ..." In this way, Varela produced a coherent theory, mathematical at base and based on fundamental notions of distinction and autonomy. This theory has enormous reach and we are only at the threshold of beginning to appreciate it and understand that his conversation is our conversation in an exchange without end, in the wholeness of our conversational domain.

#### References

L.H. Kauffman (1978). Network Synthesis and Varela's Calculus, *International Journal of General Systems* 4,(1978), 179-187.

L.H. Kauffman (1978). *DeMorgan Algebras - Completeness and Recursion. Proceedings of the Eighth International Conference on Multiple Valued Logic* (1978), IEEE Computer Society Press, 82-86.

L.H. Kauffman and F. Varela (1980). Form dynamics. Journal of Social and Biological Structures (1980), 171-206.

L.H. Kauffman. Sign and Space (1985), In Religious Experience and Scientific Paradigms. Proceedings of the 1982 IASWR Conference, Stony Brook, New York: Institute of Advanced Study of World Religions, (1985), 118-164.

L.H. Kauffman (1987). Imaginary values in mathematical logic. Proceedings of the Seventeenth International Conference on Multiple Valued Logic, May 26-28 (1987), Boston MA, IEEE Computer Society Press, 282-289.

L.H. Kauffman (1987). Self-reference and recursive forms. Journal of Social and Biological Structures (1987), 53-72.

L.H. Kauffman (1994). Knot Logic. In Knots and Applications ed. by L. Kauffman, World Scientific Pub. (1994) pp. 1-110.

L.H. Kauffman and J. M. Flagg (2017). The Flagg Resolution. (in preparation).

L.H. Kauffman, (1995), "Virtual Logic". *Systems Research*, Vol. 13 (Festschrift for Heinz von Foerster), No. 3, pp. 283-310 (1996). <A short version appears in "10th IEEE Intl. Symposium on Intelligent Control", Adrem Inc. Pub., pp. 89-98.

L.H. Kauffman, (2001), "The mathematics of Charles Sanders Peirce", in *Cy*bernetics and Human Knowing, Volume 8, No. 1-2, pp. 79-110.

L.H. Kauffman, (2003), "Eigenforms - objects as tokens for eigenbehaviours", in *Cybernetics and Human Knowing*, Volume 10, No. 3-4, pp. 73-89.

L.H. Kauffman (2005), Eigenform, Kybernetes. The Intl J. of Systems and Cybernetics, 34, No. 1/2 (2005), Emerald Group Publishing Ltd, p. 129-150.

L.H. Kauffman (2005), Reformulating the map color theorem. *Discrete Math.* 302 (2005), no. 1-3, 145-172.

L.H. Kauffman (2009), Reflexivity and Eigenform – The Shape of Process. *Kybernetes*, Vol 4. No. 3, July 2009.

L.H. Kauffman (2010), Reflexivity, Eigenform and Foundations of Physics. In *Reflexivity, Proceedings of ANPA* 30, Arleta D. Ford, Editor, Published by ANPA, June 2010, pp. 158-222.

L.H. Kauffman (2010), Reflexivity and Foundations of Physics, In "Search for Fundamental Theory" - The VIIth Intenational Symposium Honoring French Mathematical Physicist Jean-Pierre Vigier, Imperial College, London, UK, 12-14 July 2010, edited by R. Amoroso, P. Rowlands and S. Jeffers, AIP - American Institute of Physics Pub., Melville, N.Y., pp.48-89.

L.H. Kauffman (2011), Eigenforms, discrete processes and quantum processes. in "Journal of Physics: Conference Series The open access journal for conferences EmerQuM11: Emergent Quantum Mechanics 2011" (Heinz von Foerster Congress), Vienna Austria 11-13 November 2011, edited by Gerhard Grossing, Vol. 361 (2012), IOP Publishing, 012034. http://front.math.ucdavis.edu/1109.1892.

L.H. Kauffman (2015), Iterants, Fermions and Majorana Operators. In "Unified Field Mechanics - Natural Science Beyond the Veil of Spacetime", edited by R. Amoroso, L.H. Kauffman, P. Rowlands, World Scientific Pub. Co. (2015). pp. 1-32.

L.H. Kauffman (2016). Knot logic and topological quantum computing with majorana fermions. In "*Logic and algebraic structures in quantum computing and in-formation*", Lecture Notes in Logic, J. Chubb, J. Chubb, Ali Eskandarian, and V. Harizanov, editors, 124 pages Cambridge University Press (2016).

G. Spencer-Brown (1969). *Laws of Form*. George Allen and Unwin Ltd. London (1969).

F.J. Varela (1975). A calculus for self-reference. (1975). Int. J. Gen. Systems.

H. Maturana, R. Uribe and F.J. Varela (1974), Autopoesis - the organization of living systems - its characterization and a model. *Biosystems* 5 (1974) pp. 187-196.

F.J. Varela (1979), "Principles of Biological Autonomy" North Holland Series in *General Systems Research*, G. Klir editor (1979). North Holland, New York, Oxford (1979).

F.J. Varela (2001), Of organ transplantation. *Journal of Consciousness Studies*, 8, No. 5-7 (2001), p. 259-271. <a href="http://www.oikos.org/varelafragments.htm">http://www.oikos.org/varelafragments.htm</a>>

H.R. Maturana and F.J. Varela (1987), "*The Tree of Knowledge – The Biological Roots of Human Understanding*", New Science Library (1987).

H.P. Barendregt (1985). "The Lambda Calculus Its Syntax and Semantics", North Holland (1981 and 1985).

T. Stcherbatsky (1968). "Buddhist Logic", (1958) Mouton de Gruyter.

H. von Foerster (1981). "Observing Systems", *Objects: Tokens for Eigenbehaviours*, pp. 274-285. Intersystems Publications (1981).

H. von Foerster, (1981), "Notes on an epistemology for living things", in *Observing Systems, The Systems Inquiry Series*, Intersystems Publications, pp. 258-271.

H. von Foerster, (1981), "On constructing a reality", in *Observing Systems, The Systems Inquiry Series*, Intersystems Publications, pp. 288-309.

C.S. Peirce (1933), "*Collected Papers* – II, p. 2.230-2.231, edited by Charles Hartshorne and Paul Weiss, Harvard University Press, Cambridge (1933).

Lawvere, F.W. (1970), "Introduction" in Toposes, Algebraic Geometry and Logic, Springer Lecture Notes on Mathematics Vol. 274, pp. 1-12.

McCulloch, W.S. (1965), "What is a number that a man may know it, and a man, that he may know a number?", in *Embodiments of Mind*, MIT Press, pp. 1-18.

# MICHEL BITBOL\* AND PIER LUIGI LUISI\*\*

# AUTOPOIESIS WITH OR WITHOUT COGNITION: DEFINING LIFE AT ITS EDGE

# 1. Posing the question

The theory of autopoiesis, as developed by Maturana and Varela (Varela *et al.* 1974; Maturana & Varela 1980, 1998; Maturana *et al.* 1960; Varela 2000), captures the essence of cellular life by recognizing that life is a cyclic process that produces the components that in turn self-organize in the process itself, and all within a boundary of its own making. The authors thus arrived at the definition of an autopoietic unit, as a system that is capable of self-maintenance owing to a process of components self-generation from within. This generalizes the definition of life. Systems involving RNA-DNA coding (as in *actual* cells) are no longer the only possible living entities. The important notion is that the activity leading to life is a process from within, i.e. dictated by the internal system's organization. This 'activity from within' permeates all other concepts associated to autopoiesis, like the notion of autonomy, or biological evolution, or the rules of internal closure (Varela *et al.* 1974; Maturana & Varela 1980, 1998; Maturana *et al.* 1960; Varela 2000).

 <sup>\*</sup> CNRS/ENS, Archives Husserl, 45 rue d'Ulm, 75005 Paris, France <michel.bitbol@ens.fr>

<sup>\*\*</sup> Department of Biology, University of Roma 3, Rome, Italy

<sup>\*\*\*</sup> This paper appeared in *J.R. Soc. Interface* (2004) **1**, 99-107 and is reprinted by permission of the authors.

The philosophical implications of autopoiesis have been considered in the literature (Varela 1979, 1989a; Zeleny 1977, 1997; Maturana 1987; Varela et al. 1991; Luisi 2003; Fleishacker 1988). In particular, this concept has been extended to social systems, where the term 'social autopoiesis' has been coined (Luhman 1984; Mingers 1995, 1997). It has also been pointed out, that in the theory of autopoiesis there are some uncertainties and points that necessitate deeper analysis (Luisi 2003). One of these is whether autopoiesis is the necessary and sufficient condition for life. In the early days of autopoiesis, Maturana and Varela held that this must be the case. They explicitly wrote that 'autopoiesis is necessary and sufficient to characterize the organization of living systems' (Maturana & Varela 1980, p. 82). Later they identified this organization with life itself, when they asserted that 'autopoiesis in the physical space is necessary and sufficient to characterize a system as a living system' (Maturana & Varela 1980, p. 112). Similarly, Fleischaker (1998) wrote that whatever is living must be autopoietic, and that conversely, whatever is autopoietic must be living. However, according to us, the latter statement goes too far. In this paper we would like to clarify its limits.

The other point that gives rise to some uncertainty in the primary literature of autopoiesis, is the relation between autopoiesis and cognition. This relation is central in the present paper.

In this regard, it is useful to recall that Maturana and Varela, in addition to the question 'what is the blue-print of life?', had in their agenda another important question: 'what is cognition?' (Maturana & Varela 1980, 1998; Maturana *et al.* 1960; Varela 1979, 2000). In their analysis, they pointed out an indissoluble link between being a living system and *interacting with the environment*. One particular aspect of this interaction is that all living systems owe their living status to the selection of certain chemicals from the environment. These chemicals are called 'nutrients' to denote a specific relation between them and the metabolic network that incorporates them. This process of biochemical recognition occurs via a specific *sensorium*, which in turn has been developed throughout a history of coupling interactions between autopoietic units and changing environments. The authors used the term 'cognition' for this process of biological selectivity— and they came to establish a basic equivalence between life and cognition. They claimed that there is no life without cognition, and that it is the co-emergence of the autopoietic unit and its cognitive activity that gives rise to the process of life (Maturana & Varela 1980, 1998; Maturana *et al.* 1960; Varela 1979, 2000). According to them 'Living systems are cognitive systems, and living as a process is a process of cognition' (Maturana and Varela 1980, p. 13).

Together with Maturana's and Varela's previously quoted statement, this would mean complete equivalence between three processes: autopoiesis, life and cognition. At this point, our doubt about the equivalence between autopoiesis and life can be reformulated thus: *is there a real equivalence between autopoiesis and cognition?* 

The reason why the relation between autopoiesis and cognition may give rise to some confusion can be formulated in the following way: if cognition is a primary feature of life, and if there is an equivalence between autopoiesis and life, cognition should be included explicitly in the definition of autopoiesis. True, since cognition is *prima facie* a relational feature, whereas autopoiesis is an organizational feature, this inclusion does not amount to a *mere identification*. However, autopoiesis could at least include the necessity of cognition-like relations for its own maintenance in its definition. Conversely, if (as hitherto witnessed in the literature) this is not done, in other words if cognition (a) remains excluded from the definition of autopoiesis (which focuses on internal organization and self-generation) and (b) is nevertheless construed as indispensable for life, then autopoiesis and cognition are distinct processes, and autopoiesis alone may not be sufficient for defining life.

This question appears to be timely, as another group of authors has approached the same question almost simultaneously to us (Bourgine and Stewart 2004). These authors start on a different premise, presenting a mathematical model of a three-dimensional tessellation automaton of autopoiesis, and developing different arguments from ours as far as the relation with cognition is concerned. However, the conclusion reached by Bourgine and Stewart is similar to ours, with a few interesting exceptions that are discussed later on.



Figure 1. The autopoietic self-reproduction of vesicles.

# 2. A model of autopoietic fatty acid vesicles

The present paper, instead of considering theoretical models, focuses on systems that have been considered experimental expressions of chemical autopoiesis (Bachman *et al.* 1992; Luisi 1996, 1997; Walde *et al.* 1994; Wick *et al.* 1995; Zepik *et al.* 2001). They originated from a direct interaction with Varela (see Luisi *et al.* 1996) and have the advantage of being simple laboratory systems. This way, the question of whether such autopoietic chemical systems are 'living' or not can be checked against a real, concrete case.

Let us consider more closely one of those implemented chemical autopoietic systems. A typical example is represented in figure 1.

Here, one starts from a relatively static aqueous system (vesicle) formed by the surfactant S. Then, a highly lipophilic precursor of S, indicated as S-S, binds to the boundary of the vesicle and is hydroly-

sed there yielding the very surfactant S. The vesicle grows, and eventually it divides into two or more thermodynamically more stable smaller vesicles. The more vesicles that are formed, the more S-S is bound and the more vesicles are formed, i.e. the process is auto-catalytic. Since the whole process of hydrolysis and growth takes place because of and within the boundary, the vesicle can be seen as a simple self-reproducing, autopoietic system.

In fact, such systems fulfil the three criteria of autopoiesis indicated by Varela (2000). These criteria are: (1) that the system builds its own boundaries; (2) that this construction is due to reaction(s) (activity) taking place *within* the system; (3) that it is performed through reactions determined by the system itself. It is clear that these criteria apply to the system of figure 1, albeit in the limiting case that: (i) all is taking place at its boundary; and (ii) there is no activity in the aqueous core. Indeed, these systems are the simplest possible case of experimental autopoiesis. The former three criteria have also been used by Bourgine and Stewart (2004), where they are applied to the threedimensional tessellation automaton.

While the system in figure 1 corresponds to autopoiesis in a selfreproducing mode, an autopoietic system in the homeostatic mode has also been implemented experimentally (Zepik *et al.* 2001). This is characterized by two competing processes, one that forms the vesicles, the other that destroys them, both taking place at the boundary. By changing the conditions, the relative velocity of the two processes changes, and accordingly the system can be in homeostasis, can grow or can decay and 'die' (Zepik *et al.* 2001).

Now, all this represents a vesicle that adsorbs chemicals and by doing so is capable of self-generation from within, either in the mode of homeostasis or in the mode of growth and self-reproduction. What is then the difference with a bacterium, that absorbs sugar as a nutrient from the environment? At first sight, this vesicular process and the bacterial process of glucose assimilation are similar. However, we would commonly ascribe the definition of living to the second case, and generally not to the first case. Admittedly, both systems are autopoietic (Luisi 1996), but we cannot help thinking that there must be a difference between them. To substantiate the notion of this difference, we characterize it from a 'cognitive' point of view, according to an argument which is developed in the following sections.

# 3. Cognition: a non-representationalist definition, with qualifications

The first general problem is to provide a definition of cognition that is both comprehensive enough to avoid mere identification with human brain's functioning, and specific enough not to encompass any self-catalytic chemical process whatsoever. Some of these considerations have been presented before in a preliminary form (Bitbol 2001).

Maturana and Varela's theory of cognition is certainly the most radical attempt in this direction. In this theory, the relevant concept is not information provided by the external world, but local environmental conditions for maintaining an operationally closed, autopoietic *unit.* The invariants of this type of unit are said not to represent any feature of the world, but rather to identify with steady aspects of its own internal dynamical organization. As for the advisable changes of an operationally closed unit, they do not prove that the unit possesses a faithful picture of the world according to which the changes are determined, but only that its internal working is viable in relation to environmental disturbances. In other terms: cognition is definitely not tantamount to a passive reproduction of some external reality. It is instead mostly governed by the activity of the cognitive system itself. To understand this, one must realize that it is the cognitive structure that selects, and retroactively alters, the *stimuli* to which it is sensitive. By this combination of choice and feed-back, the organic structure determines (in a way *moulds*) its own specific environment; and the environment in turn brings the cognitive organization to its full development. The system and the environment make one another: cognition according to Maturana and Varela is a process of co-emergence.

True, the verbal separation between the world and the inner organization of the unit sounds like a false dichotomy *in view of the very theory of cognition that uses it.* How can someone refer to something such as an 'external independent world' if all they can say about it uses the mediation of their 'inner' categories? And, conversely, how can one claim that these categories are 'only internal', since this presupposes a contrast with the 'external independent world'? Should we then accept that the invariants of an operationally closed unit are indeed equivalent to representations, if we make clear that what they 'represent' are features of the environment that are salient relative to them, thereby carefully avoiding the slippery notion of an 'external independent world'? Should we rehabilitate the term 'representation' that was initially banished, provided we keep the former qualification in mind? These radical questions are perfectly sound; however, by implementing them too thoroughly, we run the risk of wiping out, in the vocabulary we use, the momentous difference between naive realism and Maturana and Varela's theory of cognition. Even though this difference cannot be accurately expressed by any lexical remnant of the dualist picture, it retains a *function* that we make readily accessible in the rest of this paper by using a systematically altered vocabulary (for instance, we will replace 'representation' with expressions such as 'representation-like behaviour' or 'representation-like organization' whenever necessary).

One important function of Maturana and Varela's theory is that it forces one to redefine the 'cognitive domain' of the operationally closed unit and to take advantage of this to discover new modes of cognition. Although the theory is expressed in a language that still borrows something from the representationalist paradigm it tends to replace, it retains the value of a guiding thread towards hitherto ignored (or minimized) aspects of cognition. As we mentioned at the beginning of this section, in Maturana and Varela's theory of cognition, the 'cognitive domain' is said to be no longer some fraction of a pre-existing world, but a region of the environment that has *co-evolved* with the closed unit and in which the latter's organization may persist, develop and reproduce despite the disturbances.<sup>1</sup> From this remark (and irrespective of the persistently dualist undertones of its statement) one is led to contest the universal validity of the 'view from nowhere' theory of cognition and to complement it with the idea of 'situated' or 'embedded' modes of cognition that has proved so fruitful (Clark 1997; Varela 1994). This is by itself a momentous result.

### 4. Metabolism as autopoietic cognition

Now, the next step consists of going from these very general considerations to the practical case of biological and chemical systems. Clearly, all that has just been said about cognition can be abstracted from the notion of metabolism. When an amoeba or any other living cell chooses the metabolites from the environment and rejects catabolites in it, this corresponds to a dynamic interaction that permits the enacting and the coming to being of both the living organism and the environment. In other words, metabolism always involves a dynamic interaction with the outer medium. Therefore, metabolism is already by itself the biological correlate of the notion of cognition. In this sense, our view is slightly different from the predicament of Bourgine and Stewart (2004), who write 'autopoiesis focuses naturally on the internal functioning of the organism, notably its metabolism; cognition naturally thematizes the interactions between an organism and its environment'. We believe in fact that metabolism is not only a property of the interior of the living organism. Metabolism cannot exist permanently without (mutual) interaction with the environment. In this active interaction, the organism selects its material, and in this sense a fullblown metabolism is tantamount to cognition.

A closer examination, however, shows that there are two levels of metabolism—and therefore of cognition. Firstly, the normal, steady metabolism described above usually concerns compounds that are already 'familiar', i.e. metabolites that have accompanied the life of the cell and its progeny for generations by recursive series of interactions. Secondly, in addition to this 'familiar' aspect of metabolism, there is another, albeit less frequent level that refers to the interaction with entirely novel compounds. This is important, as it opens the possibility of temporary or permanent reshuffling of the metabolic pathways of the autopoietic unit, and is associated with adaptation and evolution.

In other, more detailed, words, we should consider two aspects of metabolism/cognition.

- (1) The ordinary homeostatic metabolism that corresponds to the normal life and self-maintenance of the cell. There, the cell uses a multiplicity of standard 'nutrients' that may or may not all be present at the same time. This mode of functioning presupposes a (limited) range of possible changes in the structure of the cell, in order to incorporate and use the nutrients to which it is adapted whenever they appear in its environment. Interestingly, this ability may also extend to non-standard elements, if it happens that they have enough chemical features in common with the standard 'nutrients'.
- (2) An open-ended metabolism of elements that are 'novel' in the strongest sense, since they require an unprecedented rearrangement of the chemical pathways and basic constituents of the cell for their incorporation to become feasible. This type of alteration is of a higher order with respect to case (1): it shifts the whole range of possible changes in the cell structure, not only this structure itself. It is only restricted by the condition that it must remain within limits that do not impair the viability of the altered cell.

In regard to this twofold analysis of metabolism/cognition, it is useful to consider similar suggestions that were made by Maturana and Varela concerning how the interaction between autopoietic unit and environment can *change*.

The first suggestion is the crucial distinction between *structure* and *organization* of an autopoietic unit, which has only been alluded to up to this point. Structure is the set of actual relations that hold between the components of the unit. It embodies the pattern of processes that define *the specific physical realization of this unit in its present configuration*. In contrast, organization is a less constraining set (range) of actual or possible relations between types of components of the unit. It is the pattern of processes that define the unit as *an element of a class*: the class of viable members of a certain species, or the range of (possibly successive) realizations of the 'same' individual. Accordingly, when they deal with cognition, Maturana and Varela strong dynamical terms such as 'to change', 'to be deformed' or 'to be renewed', are pri-

marily applied to the *internal structure* (rather than the organization) of autopoietic processes. Maturana and Varela write: 'If a living system enters into a cognitive interaction, its internal state is changed in a manner relevant to its maintenance, and it enters into a new interaction *without loss of its identity*' (Maturana & Varela 1980, p. 13).

The second suggestion is Maturana and Varela's<sup>2</sup> reference to a concept of co-evolution, which assumes alteration of a higher order type: not a mere continuous drift within the framework of a single organization, but a sequence of sometimes discontinuous mutual alterations of both the environment and the very definition (or organization) of the autopoietic unit, followed by periods of relative stability due to mutual co-adaptation.

#### 5. Cognition, change and adaptation

We can now take on these two suggestions to build a more precise and selective set of constraints for the concept of cognition, and to this aim we embed Maturana and Varela's view within a detailed analysis due to J. Piaget (see Piaget 1967). This procedure is in harmony with the spirit of autopoiesis, since Piaget was explicitly quoted by Varela as a fellow-thinker in the domain of biology and cognition.<sup>3</sup> One difference between Varela's and Piaget's theories of cognition, however, is that Piaget essentially started from complex human cognition as a model for biologically more elementary forms of cognition, whereas Varela proceeded the other way round. Also, since Piaget deals essentially with perceptual inputs and motor outputs, his conception involves the latent presupposition that cognition mostly deals with *novel* features of the environment. This would correspond formally to the biological notion of a completely new (hitherto absent) foreign molecule interacting with the pre-existing metabolism of the living structure.

In the limits of this analogy, it is possible to derive a general scheme of cognition extrapolated from Maturana and Varela's scheme, and from Piaget's as well: a fine-tuned hybridization, rather than a literally orthodox Maturana–Varela's or Piaget's view.

60

Such a scheme starts from the consideration that Piaget decomposed cognition into two steps.

Step one corresponds to Maturana and Varela's change of internal structure (their first suggestion). In the context of human cognition, Piaget calls this first step of cognition the process of *assimilation*: incorporation of objects of the environment to the subject's pre-existing schemes of motor activity. In the case of cells, this would correspond to ordinary homeostatic metabolism, as discussed above; namely a process by which an operationally closed unit absorbs physical or chemical elements from the environment and integrates them somehow into its own inner processes, maintaining both its identity and its viability. In other terms, assimilation is a process by which the unit temporarily changes its detailed structure according to the incorporated elements without changing its global organization.

In the example of the bacterium, it is interesting to distinguish *active* and *passive* integration (while keeping in mind that active, rather than passive, integration of a molecular element is the only acceptable *analogon* of Piaget's 'assimilation').

- *Active*. The incorporated molecule *X* can immediately find its place in the metabolism as it stands, say as an intermediary step within an already existing chemical network. This is the case, for example, if a bacterium with appropriate enzymatic equipment finds ordinary lactose in its environment.
- *Passive*. The incorporated molecule X' can be a new neutral, nonnutrient molecule, for example a variant isomer of lactose. In this case, the molecule X' will remain inside the bacterium as a guest molecule for a certain period of time (and eventually be expelled).

However, as we have mentioned, it may occur that a novel chemical, which was formerly incorporated in a passive way (or even was poisonous) becomes actively integrated in the network of reactions of an autopoietic unit. *This requires an enduring modification of the very definition of the unit, involving a fraction or totality of its metabolism.* Thus, instead of remaining a neutral or threatening feature for it, the disturbance X' may become part and parcel of the altered unit, provided that the appropriate reorganization has taken place. We take such an alteration as *step two* of the selfprotecting transformation of an operationally closed unit; a step that would correspond to Maturana and Varela's discrete evolution or co-evolution (their second suggestion above).

This step two corresponds to what was called *accommodation* by Piaget, in the context of human cognition: drastic reorganization of the subject's scheme of motor activity in order to assimilate new objects.

For a bacterium or any other cell, however, we would rather call it *adaptation*. In this process, the unit transforms itself *permanently* and thereby becomes able to more efficiently assimilate the former disturbances and to remain viable even when confronted with higher concentrations of disturbing substances of the same type.

*Permanence* of the acquired transformations is the keyword of this second step. According to Piaget, in human behaviour the transformations of a genuinely 'accommodated' unit persist for some time (by way of representational or embodied 'memory') after the disturbance has disappeared; and they are ready to play their adaptive role again whenever the disturbance recurs.

In a bacterium, the change in metabolism may also be permanent, so that the living unit would be ready to cope efficiently with another disturbance of the same kind. One important way this can be done implies a permanent change in the genome of the bacterium. We must be careful, however, in this regard: a mutation is not to be seen as a local disturbance to be incorporated in the life cycle of a single bacterium, but it implies first a selection of a sub-species (one that better copes with the foreign substance X) in a large population of bacteria, followed by over-reproduction of this selected species. In contrast, the process of cognition is standardly taken to imply enduring *identity* of the cognizing unit as such. How can we reconcile this with the apparent loss of 'sameness' from one generation of bacteria to another? This reconciliation can indeed be achieved, provided we refer to what in the literature has been called the genidentity of the lineage of units; a form of identity that relies on historical continuity of the sequence of changes (see, e.g. Carnap 1967). A thorough discussion of this concept of 'lineage' and the associated difficulties of defi-

62



Figure 2. The membrane (S) is formed from B through a process characterized by a velocity  $v_{gen}$ . Then, S decays with velocity  $v_{dec}$ . The precursor metabolite A') enters from the environment. Furthermore, C decays into C' that is eventually expelled.

ning the target of selection can be found in a recent review by Mc-Mullin (2003).

### 6. Cognition in artificial 'metabolic' networks

At this point, the question may arise as to whether an artificial system may reach the stage of cognition, and therefore be called living. This is an important point, because it may suggest experiments of wet biochemistry to implement the minimal autopoietic *and* cognitive systems.

Let us sketch an example: that of an internal cycle of three components, A, B and C (figure 2). They (or some of their precursors, such as A') enter from the environment by ways and mechanisms that we do not need to describe in detail. One of these components, e.g. B, builds the membrane S with a velocity  $v_{gen}$ , and S decays with its own rate  $v_{dec}$ . Of the other two reagents, the first (C) decays into C'



Figure 3. The foreign substance X-Y enters into the autopoietic structure and is capable of being incorporated into the existing metabolism, then modifying it.

and is expelled to the outer environment; whereas the second (A) fulfils other functions. Thus, this system operates with only three core metabolites. The system is capable of building its boundary from within, and simulates several modes of existence of the living cell. In fact, when the reaction of formation of S,  $v_{gen}$  and the velocity  $v_{dec}$  of decay of S are numerically equal, the system is in homeostasis; if  $v_{gen}$ is greater than  $v_{dec}$ , the system can grow and eventually self-reproduce; and when the reverse is true, the system dies off.

One can actually conceive several other variations of this minimal metabolic system. In a real system (bacterium) there will be many more reagents and cycles, but the quality of the picture does not change in any essential way (except for some consequences of complexity that are mentioned below).

Let us now consider the case of a substance X-Y that interacts with the previously described autopoietic unit and is not necessarily recognized by its metabolic cycle. As previously mentioned (under the headings 'active' and 'passive' incorporation), this molecule can interact with the autopoietic unit in two or three different ways. It can be absorbed and parked inside the unit without being integrated, and then be eventually expelled. It can also block one of the reactions of the cycle (i.e. act as an inhibitor).<sup>4</sup> Or else, it can become part of the metabolic cycle as indicated in figure 3 (with hydrolysis of X-Y and release of Y into the environment). This latter course of events supposes either a pre-existing ability of the cycle to deal with the new substance (this is 'assimilation'), or an appropriate change of some element(s) of the cycle eventually promoting the required ability (this is 'accommodation'). In a cell, the latter case would correspond to a change of the genomic system, e.g. to the development of novel enzymes that are capable of accepting and transforming the molecule X-Yand inserting it in the cell metabolism. Note that, as a consequence of such an 'accommodation', the value of the constant  $v_{gen}$  and/or  $v_{dec}$ can be changed. In particular, these values can be modified in response to the alteration of the environment triggered by the release of metabolite Y. This corresponds quite well to the view of Bourgine and Stewart (see the conclusion below). Indeed, according to them, there is cognition whenever there is: (a) an environmental cause (here the outer molecule X-Y; (b) a resulting effect from the unit (here the release of a metabolite Y); and (c) an adaptive virtue of the effect (here, say, an increased rate of self-reproduction due to the alteration of the ratio  $v_{gen}/v_{dec}$ ).

By means of this model, we are able to visualize the minimal metabolic unit that also corresponds to the minimal level of cognition. This visualization is pragmatically important, in so far as it may suggest to the experimentalist some minimalist cell that can be fabricated in the laboratory. In fact, if we could realize a vesicular system hosting in its interior the 'simple' metabolic cycle of figures 2 and 3, and if this system were characterized by the self-maintenance of its metabolic pathway together with regeneration of the components from within the boundary *and* assimilation of components from without, then we would have realized a minimal cognitive system that is autopoietic and, therefore, according to our present thesis, we would be brought to conclude that such a system is, indeed, living. Metabolism involving a minimal ability to cognition in the sense of 'assimilation' is enough for that.

However, in the simple (first-generation) autopoietic system that has already been made (Zepik *et al.* 2001), in which one S is formed while one S is destroyed so as to self-maintain the systems' balance as illustrated in figure 1, there is no cognition in the terms explicated he-

re, not even the lower variety of cognition implied by the possibility of 'assimilation' of alternative substances within a metabolic network. Hence, we are entitled to conclude that these simple fatty acid autopoietic vesicles are definitely not 'living'.

Clearly, the above systems lend themselves to further development and increase of complexity. One decisive level of complexity, below the level of bacteria endowed with a genome, but above the level of the first- and second-generation autopoietic vesicle systems, might correspond to the case of self-organized criticality studied by Kauffmann (1995). According to Kauffmann, past a certain threshold of complexity and interconnectedness of a network of chemical reactions, autocatalysis is bound to occur. One might then make the distinction between (a) a loop of reactions which was highly unlikely to emerge spontaneously due to its low level of complexity, and (b) another loop (or rather network) of reactions so much richer than the first, that the probability of it (or its ancestor's) having emerged from an environment of the same level of complexity is close to unity. Preparing a system above this threshold would represent the third generation of artificial autopoietic units. These third-generation autopoietic units would be very likely to implement step one of cognition *spontaneously* (rather than artificially), since in this case, the probability of assimilation of new molecules within a persistently viable organization would be significantly increased (in the mode of a phase transition). Moreover, in view of their stability extending over a range of possible reorganizations, they could also implement step two of cognition.

#### 7. Broader implications and discussion

Finally, let us consider some more general implications of the scheme developed here.

One major implication is that the present synthesis drawn from Maturana and Varela and from Piaget can, in principle, be extrapolated towards the higher forms of behaviour, thus arriving at a stratified conception of cognition and its relation with life. We can present the following summary of the stratification.

66

- (0) The null stage of cognition corresponds to the case when the system of self-maintenance is either unaltered or superficially deformed by the irruption of a new environmental factor *without assimilating it*. This would correspond to the passive incorporation of the neutral molecule into an autopoietic unit, be it a vesicle system or a bacterium.
- (1) Stage one of cognition, which has been compared to Piaget's 'assimilation', involves integration of an environmental factor (obstacle or molecule) within the pre-existing processes of an autopoietic unit that is able to make use of such a factor as part of its defining network. We consider it as the very minimal condition for the concept of cognition to make sense and, accordingly, a basic condition for life. Insistence on the 'cognitive' status of the normal metabolism, which both maintains the identity of an organized unit and implies dynamical interaction with the environment, is a specificity of the present paper.
- (2) Stage two of cognition, which has been compared with Piaget's 'accommodation', implies enduring modification of the network of processes of an autopoietic unit that then becomes permanently redefined and reaches a new steady state of mutual co-adaptation with its environment. Accommodation is a form of evolution based on stable molecular or dynamic support. It may yield strongly 'anticipative' behaviour such as motricity. With its memory-like structures and adaptive features, it provides another crucial dimension to cognition. One interesting question at this point is the following. Is adaptation and the correlative mutation+evolution *also* a necessary condition for life? Adherents to the RNA-world hypothesis would probably insist on the primary value of evolution for the definition of life; however, we leave this issue open.
- (3) Stage three of cognition relies on highly complex types of accommodative changes resulting in *representation-like* types of behaviour (namely types of behaviour that evoke the use of a representation *from the standpoint of an external observer*, but that do not necessarily involve the possession by the unit of actual 'pictures' of its environment, let alone of an 'external independent world' (Clark)

1997)). This is where most thinkers would locate the emergence of cognition. In our opinion, however, the assumption that representation-like behaviour is a necessary condition for cognition is a philosophical prejudice that should rather be dismissed (Bitbol 2001).

(4) Stage four of cognition may finally involve several social aspects which transform it into genuine *knowledge*, either ascribing properties to intersubjective invariants (called 'objects') by means of *language*, or formulating mathematical counterparts to the reversible schemes of activity in which disturbances are embedded, in order to get intersubjectively shared *predictive rules*.

In considering, as in the present paper, minimal life at its edge, the upper levels of cognition (stages three and four) are clearly irrelevant. What is relevant in this case is the characterization of the most elementary stages of cognition: stage one and perhaps also stage two. Thus, as life has a stratification of complexity, so does cognition. In this way we find ourselves again close to the paradigm of Maturana and Varela, yet more fine-tuned. According to the canonical form of this paradigm, autopoiesis and cognition are exactly coextensive, and are actually two aspects of the same phenomenon-life (Varela 2000). Instead, according to the view developed in this paper, even though autopoiesis and cognition are indissolubly linked to each other, they are not identical. Autopoiesis is a pre-condition of cognition, cognition is coextensive to life, but since not every autopoietic system is thereby undergoing cognition, not every autopoietic system is a living entity.

As already mentioned, Bourgine and Stewart (2004) arrived at similar conclusions, based on an elaborate and elegant mathematical treatment. Their autopoietic three-dimensional tessellation automaton is autopoietic, but not cognitive, and therefore, they claim, not living. However, the constraints they impose on the definition of cognition are not exactly the same as ours and this yields one interesting divergence with us.

To begin with, according to these authors, interactions of a unit with an environment can be of two types:

68

- (a) inputs of the unit from the environment;
- (b) specific outputs of the unit on this environment.

Then, this twofold mode of interaction can be called 'cognition' if and only if an (a)-type interaction serves to trigger a (b)-type interaction, which promotes the viability of the system by modifying the environment in an appropriate way. Provided the latter condition is fulfilled, the (a)-mode interaction can be called 'sensation' and the (b)-type 'action'. In other words, cognition must imply active interventions on the environment in order to impose or maintain the conditions for survival.We broadly agree with this approach, provided extended homeostasis (by steady co-adaptation of the autopoietic unit and the environment, involving selective intake and outflux of molecules) is itself defined as a special variety of sensation+action. As we emphasized earlier, the discriminative use of metabolites that the living organism makes during its normal steady state cycle is indeed the most basic form of cognition, without which there would be no life.

However, there is also a small point of disagreement between us and Bourgine and Stewart. Our primary emphasis is on permanent conditions of selfmaintenance (or promotion of viability) by homeostasis, then pointing out that this involves input from and output to the environment. Bourgine's and Stewart's primary emphasis is rather directly on the input- output scheme (the sensory-motor loop), then adding a condition of cooperation of inputs and outputs for the sake of viability. This slight difference in order and emphasis has a noticeable consequence. By modifying and weakening the condition of viability, they are ready to ascribe 'cognition' to entities (such as robots) that are admittedly not autopoietic. They then add to our common statement that 'there can be autopoiesis without cognition', the further statement that 'there can be cognition without autopoiesis'. This latter statement harmonizes well with the present common wisdom according to which robots or even computers may embody (artificial) cognition. However, it does not fit with a more specific and more biological-like definition of cognition, such as ours, according to which (a) cognition is coextensive to homeostatic metabolic processes, and (b) mainly non-homeostatic contraptions such as robots or computers are cognitive tools or models, rather than entities endowed with cognition in the first place. Thus, we could say that the conception we propose in this paper is half-way between Maturana and Varela's strict equivalence of autopoiesis and cognition, and Bourgine's and Stewart's radical dissociation of autopoiesis and cognition. According to Bourgine and Stewart's is final tentative thesis, 'A system that is both autopoietic and cognitive (...) is a living system'. However, our own corresponding tentative thesis should be 'A system that is minimally cognitive and, therefore, autopoietic, is a living system'.

The authors wish to acknowledge valuable comments on a previous draft by E. Thompson, J. Stewart, S. Lazzara, I. Peschard and three anonymous referees.

#### Notes

- 1 Note that this alternative account of cognition must be made self-consistent by applying to itself. According to its own logic, it is not to be construed as a faithful picture of cognitive processes but only as a viable, efficient, fruitful way of dealing with cognition.
- 2 'If one may consider the environment of a system as a structurally plastic system, the system and its environment must then be located in the intricate history of their structural transformations, *where each one selects the trajectory of the other one*' (Varela 1989b). On the concept of 'structural coupling', the main sources are Maturana & Varela (1980, 1986).
- 3 '(. . . ) The Piagetian perspective of biological assimilation can be rephrased very naturally in the context presented here of autonomous systems and structural plasticity' (Varela 1979, p. 256).
- 4 Certain substances may even poison the unit, however a poisoning effect always corresponds to a specific interaction with the metabolic pathway, for example the inhibition of some enzymes. In other words, the poisoning corresponds to assimilation, which in this case would have a deleterious effect.

#### References

- Bachman, P., Luisi, P.L. & Lang, J. 1992 Autocatalytic selfreplicating micelles as models for prebiotic structures. *Nature* 357, 57–59.
- Bitbol, M. 2001 Non-representationalist theories of knowledge and quantum mechanics, SATS (Nordic Journal of Philosophy) 2, 37–61. Available at: http://philsciarchive.pitt.edu/archive/00000888/index.html.

70

Bourgine, P. & Stewart, J. 2004 Autopoiesis and cognition. Artificial Life 10, 327-345.

- Carnap, R. 1967 *The logical structure of the world and pseudoproblems in philosophy* (translated by R. A. George), p. 199. Berkeley, CA: University of California Press.
- Clark, A. 1997 Being there. Cambridge, MA: MIT Press.
- Fleischacker, G. 1988 Autopoiesis: the status of its system logic. *Biosystems* 22, 37-49.
- Kauffmann, S. 1995 'What is life?': was Schr"odinger right? In *What is life, the next fifty years* (ed. M. Murphy & L. A. J. O'Neill). Cambridge: Cambridge University Press.
- Luhman, K. 1984 Soziale systeme. Frankfurt: Suhrkamp.
- Luisi, P.L. 1996 Self-reproduction of micelles and vesicles: models for the mechanisms of life from the perspective of compartmented chemistry. *Adv. Chem. Phys.* 92, 425–438.
- Luisi, P.L. 1997 Self-reproduction of chemical structures and the question of the transition to life. In Astronomical and biochemical origins and the search for life in the universe (ed. C. B. Cosmovici, S. Bowyer & D. Westhimer), pp. 461–468. Bologna: Editrice Compositori.
- Luisi, P.L. 2003 Autopoiesis: a review and reappraisal. *Naturwissenschaften* 90, 49-59.
- Luisi, L., Lazcano, A. & Varela, F. 1996 Autopoiesis: the very idea. In *Defining life: the central problem in theoretical biology* (ed. M. Rizzotti), pp. 146–167. Padova: Università di Padova.
- Maturana, H. 1987 The biological foundations of consciousness. American society for cybernetics conference workbook: texts in cybernetics. Felton, CA: American Society for Cybernetics.
- Maturana, H., Lettvin, J., McCulloch, W. & Pitts, W. 1960 Life and cognition. *Gen. Physiol.* 43, 129–175.
- Maturana, H. & Varela, F. 1980 Autopoiesis and cognition: the realization of the living. Boston, MA: Reidel.
- Maturana, H. & Varela, F. 1986 *The tree of knowledge: the biological roots of human understanding.* Boston, MA: New Science Library.
- Maturana, H. & Varela, F. 1998 The tree of knowledge. Boston, MA: Shambala.
- McMullin, B. 2003 Replicators don't. Available at: http://www.eeng.dcu.ie/!autonomy/ecal95/rpl-l2h/rpll2h.html.
- Mingers, J. 1995 Self-producing systems: implications and applications of autopoiesis. New York: Plenum.

- Mingers, J. 1997 A critical evaluation of Maturana's constructivist family therapy. Syst. Practice 10(2), 137–151.
- Piaget, J. 1967 Biologie et connaissance. Paris: Gallimard.
- Varela, F. 1979 Principles of biological autonomy. NewYork: North-Holland/Elsevier.
- Varela, F. 1989a Reflections on the circulation of concepts between a biology of cognition and systemic family therapy. *Family Process* 28, 15–24.
- Varela, F. 1989b Autonomie et connaissance, p. 64. Seuil.
- Varela, F. 1994 Autopoiesis and a biology of intentionality. In *Autopoiesis and perception* (ed. B. McMullin & N. Murphy). Dublin: Dublin City University Press.
- Varela, F. 2000 El fenómeno vida. Santiago, Chile: Dolmen Esayo.
- Varela, F., Maturana, H. & Uribe, R. 1974 Autopoiesis: the organization of living systems, its characterization and a model. *Biosystems* 5, 187–195.
- Varela, F., Thompson, E. & Rosch, E. 1991 *The embodied mind: cognitive science and human experience.* Cambridge, MA: MIT Press.
- Walde, P., Wick, R., Fresta, M., Mangone, A. & Luisi, P. L. 1994 Autopoietic self-reproduction of fatty acid vesicles. J. Am. Chem. Soc. 116, 11 649–11 654.
- Wick, R., Walde, P. & Luisi, P. L. 1995 Autocatalytic selfreproduction of giant vesicles. J. Am. Chem. Soc. 117, 1435–1436.
- Zeleny, M. 1977 Self-organization of living systems formal model of autopoiesis. Int. J. Gen. Syst. 4, 13–28.
- Zeleny, M. (ed.) 1997 Autopoiesis: a theory of the living organization. New York: North-Holland.
- Zepik, H.H., Bloechliger, E. & Luisi, P. L. 2001 *A chemical model of homeostasis*. Angew. Chem. 40, 199–202.
## ARTURO CARSETTI

# THE EMERGENCE OF THE ENACTIVE MIND IN AMBIENT MEANING

Starting from the revolution brought on by J. Monod at the level of molecular Biology also by means of the graft he operated on the body of this field of research of the methods and instruments deriving from classical Cybernetics and Shannon's Information Theory, the great issues related to circularity and feed-back procedures as they are given at the level of the Bios, have come to acquire, as we have seen, an ever greater importance, thus opening up to the construction of renewed models with a view to achieving a deeper understanding of the mechanisms of life.

From Monod we passed on to Jacob and the interaction between the possible and the actual to finally arrive, at the end of a wild cavalcade that touched on all topics related to contemporary molecular Biology, to the great issues related to the alternative splicing, the ENCODE Project, the synthetic life as outlined by Craig Venter, the CRISP methodology etc. It is necessary, however, to point out that some of the issues that characterized the above-mentioned cavalcade on the more strictly biological and experimental level, such as, for instance, the issues concerning the circularity of life, the process of Self-organization and so on, were already, in the early seventies, to the attention of many scholars who - in comparison to the biologists more directly engaged on the experimental front - worked as scientists "next door" proving to be, in any case, able to make further and important contributions, above all at the

theoretical level, to the development of molecular Biology. These contributions have come to open new horizons with respect to the traditional formulations of the concepts of information and complexity thus allowing to investigate more deeply the self-organization phenomena characterizing the development of the Bios. In particular, on the basis of these contributions it was, therefore, possible to correlate the model initially introduced by Monod with the methods and the principles proper to the second-order Cybernetics. Among these strands, an important role is certainly still played by the doctrine of Autopoiesis as it has been outlined since the seventies by H. Maturana and F. Varela.

According to Maturana and Varela circularity is at the base of the biological processes. Their attempt to frame this circularity on a theoretical level is based to a large extent on the adoption of specific quantitative and informational tools. In particular, their doctrine has enlightened some of the pregnant aspects of this circularity both in relation to the construction of the Self and the outlining of the membrane. What is, however, even more important is the fact that the two scholars came to give a first mathematical characterization of the aforementioned circularity by recourse to the mathematics of self-reference and self-organizing circuits; hence the engagement of their doctrine in the wider fields of Reflexive domains, Denotational Semantics, second-order Cybernetics etc. Starting from the original papers published by Maturana and Varela, the seventies and the eighties saw the blossoming of a series of very important researches ranging from Maturana and F. Varela to S. Kauffman, H. von Foerster, L. Kauffman, A. Carsetti etc. In this context, the path of research pursued by F. Varela gradually took on an increasingly central role, as rich as it was of fruitful suggestions. The early eighties mark, in particular, a turning point quite important to what concerns such a path, a path that had seen the great scholar primarily engaged in the previous decade in the definition of the mathematical foundations of the doctrine of Autopoiesis. This turning point is consumed in large part when Varela is moving to Europe in order to continue and deepen his research. It is in the eighties, in fact, that some important papers by Varela are published just as Varela flies several times from the Americas to France and Italy as visiting professor. In 1984, in particular, a special issue of the Journal La Nuova Critica (the Italian Journal for the Philosophy of Science whose Scientific Committee he was called

to take part at the invitation of V. Tonini) entitled : "Autopoiesi e teoria dei sistemi viventi" was published in Italy in the framework of his work as visiting professor at the University of Rome and under his direct supervision.<sup>(1)</sup> This issue of *La Nuova Critica* opens with an article by Varela expressly translated by him from French into English on the occasion of the publication of this special issue of the Journal. At the heart of this choice, there was the will both to submit a paper that appeared to be a turning point in the scientific life of Varela, to the attention of a wider audience of scholars (thus providing, in particular, an adequate summary of the versatility of the research carried out by the great scholar in the years that preceded his move to Europe) and to insert it in the context of a collection of papers devoted to the doctrine of Autopoiesis also in view of a critical comparison with the theoretical approach provided at the computational and mathematical level by Denotational semantics.

For Varela, the issue represented the way both to be confronted with the Denotational Semantics as outlined by D. Scott (and to which a specific attention had been dedicated by A. Carsetti in the article entitled: "Semantica denotazionale e sistemi autopietici" included in the aforementioned issue of *La Nuova Critica* on the basis of what has been agreed by the Editor of the Journal with Varela) and to develop his research in the field of the mathematics of Self-reference but in unitary connection with Brouwer's fixed point theorem.

As is well known, the scientific vision of Varela was profoundly influenced by the analysis of the volume by Spencer Brown concerning the Laws of the Form published in 1969.<sup>(2)</sup> It is precisely the common interest in this volume which had, on the other hand, constituted the trait-d'union that initially had come to tie the Chilean scholar to the great American mathematician Louis Kauffman. As Louis Kauffman remarks: " Prior to that I had written a paper using Francisco's "Calculus for Self- Reference" to analyze the temporal behavior of self-referential circuits [L. H. Kauffman (1978) – Network Synthesis]. My papers were inspired by Varela's use of the reentering mark in his analysis of the completeness of the calculus for self-reference that he associated with that symbol. I also started corresponding with Francisco., telling him all sorts of ideas and recreations related to self-reference. We agreed to meet, and I visited him in Boulder, Colorado in 1977. There we made a plan for a paper using the waveform arithmetic. This became the paper "Form Dynamics", eventually published in the Journal for Social and Biological Structures [L. H. Kauffman and F. Varela (1980)]. Francisco based a chapter of his book "Principles of Biological Autonomy" on form dynamics. I remember being surprised to find some of my words and phrases in the pages of his book. The point about Form Dynamics was to extend the notion of autonomy inherent in a timeless representation of the reentering mark to a larger context that includes temporality and the way that time can be implicit in a spatial or symbolic form. Thus the reentering mark itself is beyond duality, but implicate within it are all sorts and forms of duality from the duality of space and time to the duality of temporal forms shifted in time from one another, to the duality of form and nothingness "itself". I believe that both Francisco and I felt that in developing Form Dynamics we had reached a balance in relation to these dualities that was quite fruitful, creative and meditative. It was a wonderful aesthetic excursion into basic science. This work relates at an abstract level with the notions of autonomy and autopoiesis inherent in the earlier work of Maturana, Uribe and Varela [H. Maturana, R. Uribe and F. J. Varela (1974)]. There they gave a generalized definition of life (autopoiesis) and showed how a self-distinguishing system could arise from a substrate of "chemical" interaction rules. I am sure that the relationship between the concept of the reentering mark and the details of this earlier model was instrumental in getting Francisco to think deeply about Laws of Form and to focus on the Calculus for Self-Reference. Later developments in fractal explorations and artificial life and autopoesis enrich the context of Form Dynamics. At the time (around 1980) that Francisco and I discussed Form Dynamics we were concerned with providing a flexible framework within which one could have the "eigenforms" of Heinz von Foerster [Heinz von Foerster (1981)] and also the dynamical evolution of these forms as demanded by biology and by mathematics. It was clear to me that Francisco had a deep intuition about the role of these eigenforms in the organizational structure of biology. This is an intuition that comes forth in his books."(3)

Following his move to France and the work done at the CREA, Varela came to develop a body of work that still leaves us amazed. In

particular, the research he carried out in the fields concerning not only the mathematics of self-reference but also the nature of the enactive mind appears very incisive. These areas of research, starting from the first setting offered by the Chilean scholar, came soon to be part of the territories of election of a multiplicity of researchers working in the field of cognitive sciences. The conviction that since the seventies animates the analysis carried forward by Varela in this areas of research is linked, first of all, to the intuition on his part of the indissoluble connection existing between life and cognition, a connection that, in his opinion, is at the basis of the same circularity of life. It is this conviction that binds him deeply to the studies carried out by H. von Foerster and Louis Kauffman and it is precisely this same basic conviction that will lead Varela to confront, in a renewed way, when necessary, the problems related to the relationship between cognition and reality both from a biological point of view and from a cognitive point of view A striking testimony of the importance of Varela's work in this field of research is represented by the fact that Varela's intuitions concerning the link between Enactivism and Realism, (intuitions that were also born on the base of some deep theoretical contributions by Prigogine as well as of some more ancient observations by von Foerster) now come to reveal their convergence, albeit limited and purely speculative, with some of the theses of the so-called participatory Realism as outlined\_by some brilliant theoretical physicists who work at level of Quantum Mechanics (cf. for instance Christopher Fuchs) but in the wake of some primitive intuitions by B. de Finetti : "La prévision: ses lois logiques, ses sources subiectives", Ann. Inst. H. Poincaré, 7: 1-68.(4)

To fully understand all the fecundity and the amplitude of the horizons which will gradually be opened at the international level starting from the moment of the first unfolding of Varela's work at the CREA and to come to understand the importance of such work for the development of the studies concerning the individuation of specific mathematical methods able to shed light on the nature of mind, we need to take care of the paper written by Varela in collaboration with Andrade (Cf. J. Andrade, and F. Varela (1984), "Self-reference and fixed points", *Acta Applic.Matem*, 2:1-19). As we have just said, following the studies carried out in America on the basis of the fundamental contributions by Spencer Brown as well as in the footsteps of the primitive intuitions by

H. von Foerster (as revisited by L. Kauffman), Varela at the time of his transfer to Europe felt the need to put into safety, so to speak, the work previously carried out at the theoretical level in the field of the mathematics of self-reference and self-organizing circuits. This indeed was the only route available to him with a view to adequately open his mind to the new world represented by Phenomenology. Only once had he reached this goal, did he fell able to deal with the great themes of Reflexivity albeit under the perspective of the embodied cognition and Enactivism. This in turn will open to the possibility of revisiting the original doctrine of Autopoiesis in the light of the studies carried out in the research field of Neurophenomenology. Hence the importance in the eyes of Varela of the paper written with Andrade, a paper intended to insert, as we have already mentioned, the reflexive domains in the unitary frame offered by Brouwer's theorem.

In this way, not only was he able to come into contact even more incisively with Denotational Semantics which had arisen since the seventies at the hands of D. Scott, but also to find, on a strictly formal level, precise cues for what will be, then, the continuation of his research concerning the role played by the recursive processes at the level of the constitution of the eigenforms.

The CREA soon became a popular destination for many scholars from America who felt the need to come to deal critically with the new climate of thought that was coming to birth. First of all D. Dennett, the Dennett who, as a proud champion of reductionism, comes to Paris as visiting professor at the CREA to deepen his research on the nature of mind in accordance with a close confrontation with both Varela and, ideally, M. Merleau Ponty. But next to Dennett and other well-known scholars there were also many young American PhD, as, for example, Cristine Skarda, who, on the instructions of their teachers, came to visit the CREA with great joy of Varela and with the birth also of triangulations of studies between La California, Paris, Brussels and Rome. (cf. the issue of La Nuova Critica. Nuova Serie N. 18). In this context, the volume by F. Varela, E. Thompson and E. Rosch The Embodied Mind (1991, The MIT Press, London), is of paramount importance. The volume was preceded by two major papers: J. Soto-Andrade and F.Varela (1990), "On mental rotations and cortical activity patterns: A linear representation is still wanted", *Biological Cybernetics*, 64:221-223 e F. Varela (1990)," Between Turing and quantum mechanics there is a body to be found", *Beh.Brain Sci. (Commentary)* 13: 687-688. In these two papers Varela not only explores the mathematics of self-reference but also addresses his attention to the theoretical principles of the second-order Cybernetics in the footsteps of von Foerster.

As Varela writes in the Introduction to the volume : "We like to consider our journey in this book as a modern continuation of a program of research founded over a generation ago by the French philosopher, Maurice Merleau-Ponty...... We hold with Merleau-Ponty that Western scientific culture requires that we see our bodies both as physical structures and as lived, experiential structures-in short, as both "outer" and "inner," biological and phenomenological. These two sides of embodiment are obviously not opposed. Instead, we continuously circulate back and forth between them. Merleau-Ponty recognized that we cannot understand this circulation without a detailed investigation of its fundamental axis, namely, the embodiment of knowledge, cognition, and experience. For Merleau-Ponty, as for us, embodiment has this double sense: it encompasses both the body as a lived, experiential structure and the body as the context or milieu of cognitive mechanisms."<sup>(5)</sup>

Varela assumes that the cognitive subject alias the Self is fundamentally fragmented, divided, or non unified It is with reference to such a conception that Varela will then introduce the definition of what is enactive: " In the enactive program, we explicitly call into question the assumption-prevalent throughout cognitive science that cognition consists of the representation of a world that is independent of our perceptual and cognitive capacities by a cognitive system that exists independent of the world. We outline instead a view of cognition as embodied action and so recover the idea of embodiment that we invoked above. We also situate this view of cognition within the context of evolutionary theory by arguing that evolution consists not in optimal adaptation but rather in what we call natural drift. This fourth step in our book may be the most creative contribution we have to offer to contemporary cognitive science."<sup>(6)</sup>

In this context, Varela is to dive in the words by Merlau Ponty:

"When I begin to reflect, my reflection bears upon an unreflective experience, moreover my reflection cannot be unaware of itself as an event, and so it appears to itself in the light of a truly creative act, of a changed structure of consciousness, and yet it has to recognize, as having priority over its own operations, the world which is given to the subject because the subject is given to himself .... Perception is not a science of the world, it is not even an act, a deliberate taking up of a position; it is the background from which all acts stand out, and is presupposed by them: The world is not an object such that I have in my possession the law of its making; it is the natural setting of, and field for, all my thoughts and all my explicit perceptions." As Varela remarks: "Mind awaken in a world. We awoke both to ourselves and to the world we inhabit. We come to reflect on that world as we grow and live."(7) According to Merleau Ponty: The subject is inseparable from the world, but from a world which the subject itself projects. Hence the ultimate secret of that essential circularity that permeates the cognitive activity Thus we come to witness the convergence of Varela's main thesis with the new evolutionary view as advocated by D. Lewontin.

In 1993 the American Journal of Psychology published a review by Daniel C. Dennett of the volume The Embodied Mind. As the great scholar in primis remarks: " cognitive science proclaims that in one way or another our minds are computers, and this seems so mechanistic, reductionistic, .....unbiological. It leaves out emotion, or what philosophers call qualia, or value,..... It doesn't explain what minds are so much as attempt to explain minds away."<sup>(8)</sup> Dennett acknowledges the importance of the enactive proposal and the congruence between this same proposal and the general theses outlined by Lewontin but he cannot hide his puzzlement. "There is something to this, of course, - he will write more later - but just how important is it? What are the relative proportions of organismic and extra-organismic contributions to the "enacted" world? It is true, as Lewontin has often pointed out, that the chemical composition of the atmosphere is as much a product of the activity of living organisms as a precondition of their life, but it is also true that it can be safely treated as a constant (an "external", "pregiven" condition), because its changes in response to local organismic activity are usually insignificant as variables in interaction with the variables under scrutiny."<sup>(9)</sup> These words well testify as the debate between Dennett and

Varela was not based on prejudices but on the precise reference to experimental data and mathematical proofs.

As we have just said, Varela and Maturana with their doctrine of Autopoiesis had enlightened, from the outset of their analysis, some of the pregnant aspects of that particular circularity characterizing the living beings both in relation to the construction of the Self and the outlining of the membrane. After the publication of the book *The Embodied Mind* in 1991, J. Brockman edited in 1995 a collective volume entitled *The Third Culture* in which are present essays by Varela, Dennett, Kauffman etc. It is precisely in his contribution to the volume that Varela comes to delineate in a perhaps more incisive way both the sense of circularity that is given at the biological level and the sense of that circularity which is given, in its turn, at the cognitive level, thus arriving to identify their indissoluble unity. The creative circle comes to close but at the same time the identity of a new philosophical doctrine achieves its most appropriate image.

As Varela remarks: "Autopoiesis attempts to define the uniqueness of the emergence that produces life in its fundamental cellular form. It's specific to the cellular level. There's a circular or network process that engenders a paradox: a self-organizing network of biochemical reactions produces molecules, which do something specific and unique: they create a boundary, a membrane, which constrains the network that has produced the constituents of the membrane. This is a logical bootstrap, a loop: a network produces entities that create a boundary, which constrains the network that produced the boundary. This bootstrap is precisely what's unique about cells. A self-distinguishing entity exists when the bootstrap is completed. This entity has produced its own boundary. It doesn't require an external agent to notice it, or to say, "I'm here." It is, by itself, a self- distinction. It bootstraps itself out of a soup of chemistry and physics.....In order to deal with the circular nature of the autopoiesis idea, I developed some bits of mathematics of self-reference, in an attempt to make sense out of the bootstrap — the entity that produces its own boundary. The mathematics of self-reference involves creating formalisms to reflect the strange situation in which something produces A, which produces B, which produces A. That was 1974. Today, many colleagues call such ideas part of complexity theory. The more re-

cent wave of work in complexity illuminates my bootstrap idea, in that it's a nice way of talking about this funny, screwy logic where the snake bites its own tail and you can't discern a beginning. Forget the idea of a black box with inputs and outputs. Think in terms of loops. My early work on self-reference and autopoiesis followed from ideas developed by cyberneticists such as Warren McCulloch and Norbert Wiener, who were the first scientists to think in those terms. But early cybernetics is essentially concerned with feedback circuits, and the early cyberneticists fell short of recognizing the importance of circularity in the constitution of an identity. Their loops are still inside an input/output box. In several contemporary complex systems, the inputs and outputs are completely dependent on interactions within the system, and their richness comes from their internal connectedness. Give up the boxes, and work with the entire loopiness of the thing. For instance, it's impossible to build a nervous system that has very clear inputs and outputs."(10) Hence the almost immediate idea of applying the logic of emergent properties proper to circular structures in order to investigate the functioning of the nervous system. Another scientific revolution enters the scene: a revolution focused on the analysis of the self-organization processes that inhabit the most hidden recesses of brain functioning.

"The consequence is a radical change in the received view of the brain. The nervous system is not an information-processing system, because, by definition, information-processing systems need clear inputs. The nervous system has internal, or operational, closure. The key question is how, on the basis of its ongoing internal dynamics, the brain configures or constitutes relevance from otherwise non meaningful interactions. You can see why I'm not really interested in the classical artificialintelligence and information-processing metaphors of brain studies. The brain can't be understood as a computer, in any interesting sense, and I part company with the people who think that the brain does rely on symbolic representation. The same intuitions cut across other biological fields. Deconstruct the notion that the brain is processing information and making a representation of the world. Deconstruct the militaristic notion that the immune system is about defense and looking out for invaders. Deconstruct the notion that evolution is about optimizing fitness to live in the conditions present in some kind of niche. I haven't been directly active in this last line of research, but it's of great importance for my argument. Deconstructing adaptation means deconstructing neo-Darwinism. Steve Gould, Stuart Kauffman, and Dick Lewontin, each in his own way, have spelled out this new evolutionary view. Lewontin, in particular, has much appreciated the fact that my work on the nervous system mirrors his work with evolution."<sup>(11)</sup>

To the volume *The Third Culture* will then follow in 1999 another volume entitled : *Naturalizing Phenomenology. Issues in Contemporary Phenomenology and Cognitive Science*, edited by Jean Petitot, Francisco J. Varela, Bernard Pachoud, and Jean-Michel Roy, a volume that sees the presence of many qualified scholars who had collaborated with the CREA such as Petitot, Roy, Thompson, Petit, Longo, Dupuy etc. The volume represents an important challenge: with its publication it is not only Varela who comes to terms with Husserl's heritage : It is a large share of the scientific French context which comes to confront this philosophical legacy.

Whilst in the last days of his life Varela continued to open up unexplored horizons at the level of an area of research that was increasingly to be characterized in cognitive terms (with special attention to the ongoing discoveries in the field of Neurophenomenology), Louis Kauffman, in the meantime, was developing, throughout his last fruitful twenty years of research, a more and more incisive deepening at the mathematical level of the issues concerning the reflexive domains (and in general the cybernetic study of Reflexivity) in agreement, in particular, to the intuitions introduced in this field of research by H. Foerster by means of the utilization of the methods proper to the second-order Cybernetics.

"If science is to be performed - he writes - in a reflexive domain, one must recognize the actions of the persons in the domain. Persons and their actions are not separate. if an action is a scientific theory about the domain, then this theory becomes a (new) transformation of the domain. In a reflexive domain, theory inevitably affects the ground that it studies. The fact that an entire reflexive domain can be seen as an eigenform suggests the observation of that domain in a wider view. For example, physics can be seen as a reflexive domain and one can take a metascientific view, allowing physics itself to be one of the objects of a larger domain in which it (physical science) is one of the eigenforms. Once cybernetics is defined in terms of itself, it becomes what is commonly

called "second-order cybernetics." From the point of view of this article, I identify cybernetics as second-order cybernetics and take it (cybernetics) to be itself a reflexive domain. in this way, cybernetics is both the origin of the concept of reflexive domain and is itself an exemplar of that concept. in George Spencer Brown's book Laws of Form (Spencer Brown 1969) a very simple mathematical system is constructed on the basis of a single sign, the mark, designated by a circle or a box or a rightangle bracket. i shall not develop the formalism here. We can take < > to stand for the Spencer Brown mark. But that very sign, < >, in our eyes, makes a distinction in the plane in which it is drawn, and the interpretation of Spencer Brown's calculus has us understand that the sign refers to the distinction that the sign makes (we make that distinction when we are identified with the sign). Thus, the sign of distinction in the calculus of Spencer Brown is self-referential. The equation  $\langle \rangle = \langle \rangle$  can be understood as an eigenform equation. We interpret the mark on the left as a transformation from the void of its inside to the marked state that is seen on its outside. We interpret the mark on the right as a mark of distinction. The identity of the mark of distinction with the act of distinguishing is the fundamental eigenform of Laws of Form. it is a conceptual fixed point, not a notational one, and this means that there is no excursion to infinity in this eigenform. The sign indicates the very distinction that the sign makes. Everything is said in the context of an observer. The observer herself makes the distinction that is its own sign. in the form, the sign, the first distinction and the observer are identical. The key point about the reflexivity of the sign of distinction in Laws of Form is that it does not lead to an infinite regress. it is through concept that our own thought is kept from spiraling into infinite repetition. Eigenform occurs at the point that infinite repetition is replaced by fixed point and by concept. in this way, our perception is always a precise mixture of sense data and the sense of thought".(12) Hence a deep relationship with the Fixed Point Theorem (that tells that every A in a reflexive domain has an eigenform).

As Louis Kauffman remarks in another paper: " In fact, ultimately, the form of an "object" is the form of the distinction that "it" makes in the space of our perception. In any attempt to speak absolutely about the nature of form we take the form of distinction for the form (paraphrasing Spencer-Brown 1969). It is the form of distinction that remains constant and produces an apparent object for the observer. How can you write an equation for this? The simplest route is to write O(A)=A. The object A is a fixed point for the observer O. The object is an eigenform. We must emphasize that this is the most schematically possible description of the condition of the observer in relation to an object A. We only record that the observer as an actor (operator) manages through his acting to leave the (form of) the object unchanged. This can be a recognition of the symmetry of the object but it also can be a description of how the observer, searching for an object, makes that object up (like a good fairy tale) from the very ingredients that are the observer herself. This is the situation that Heinz von Foerster has been most interested in studying. As he puts it, if you give a person an undecideable problem, then the answer that he gives you is a description of himself. And so, by working on hard and undecideable problems we go deeply into the discovery of who we really are. All this is symbolized in the little equation O(A)=A. We can start anew from the dictum that the perceiver and the perceived arise together in the condition of observation. This is a stance that insists on mutuality (neither perceiver nor the perceived causes the other). A distinction has emerged and with it a world with an observer and an observed. The distinction is itself an eigenform We identify the world in terms of how we shape it. We shape the world in response to how it changes us. We change the world and the world changes us. Objects arise as tokens of a behavior that leads to seemingly unchanging forms. Forms are seen to be unchanging through their invariance under our attempts to change, to shape them".<sup>(13)</sup>

Thus, in accordance with Kauffman's main thesis, the notion of a fixed object becomes a notion concerning the process that produces the apparent stability of the same object. This process can be simplified in a model to become a recursive process where a rule or rules are applied time and time again. The resulting object of such a process is the eigenform of the process, and the process itself is the eigenbehavior. In other words, we have a model for thinking about objects as tokens for eigenbehavior as advocated by H. von Foerster. In particular, Kauffman's model examines the result of a simple recursive process carried to its limit. For example, suppose that X = (F(F(F...))), that is, each step in the process enclose the results of the previous steps within a frame. These objects, these infinite nest of frames, may go beyond the specific proper-

ties of the world in which we operate. They attain their stability through the limiting process that goes outside the immediate world of individual actions. We need, in this sense, an imaginative act to complete such objects to become tokens for eigenbehaviors. It is impossible to make an infinite nest of frames: as the great scholar remarks, we do not make it, we imagine it. And in imagining that infinite nest of frames, we arrive at the eigenform. "The leap of imagination to the infinite eigenform is a model of the human ability to create signs and symbols. In the case of the eigenform X with X = F(X), X can be regarded as the name of the process itself or as the name of the limiting process."(14) It is in the exercise of his imagination that Narcissus will be able to come into contact with the existence of his I until he himself is reflected and objectified in it at the level of the resulting tissue of the involved eigenforms. The image marks his advent to the truth and it is in this same image that Narcissus drowns. In this sense, the form of an object is the form of the distinction that it operates in the space of perception.

At the level of the ancient Myth, the Minotaur is initially pure magma (and as such he is blind, pure plot of intensities devoid of boundaries in himself). When the distinction is born (as eigenform) and with it the object, this means that the form of the distinction (which remains constant) has produced an object for the observer. The Minotaur succeeds in opening his eyes in the same time that an object comes to appear. Unity of the structures of perception, on the one hand, and of the objects seen, on the other hand. Reality brings into play a form of distinction that remaining constant produces, in turn, an object for the observer-operator: O(A) = A, where the object A constitutes an eigenform for the observer O. When the distinction emerges here is both perceiver and perceived, and here is a particular ability that draws existence along its own realizing itself according to a form in action. In this context, the Minotaur appears as a paradigmatic example of a non-trivial machine in the sense of von Foerster, i.e. of a self-organizing machine. In other words, at the level of the Minotaur, an eigenform constitutes the order for the process that generates it. Always according to the ancient Myth, Narcissus comes to existence by drowning in his image as a fixed point of self-replication. From a general point of view, the object is given by the process that determines the stability (invariance) of the object and this process is recursive. It is, however, every time, a certain type

of self-organization that gives rise to the specific embodiment related to this type of process. This embodiment would not, in fact, be possible (as we have just seen) without the enchantment created by the flute played by Pan: the universe of intensities necessarily follows the experimentum and the "irruption". Starting from Narcissus, moreover, there is no novelty that does not constitute a variation that arises from a Chance that represents, in turn, the other face of the Empire of Necessity. Here is the eternal spinning of the Wheel of Time in order to ensure through Omega (in accordance with Chaitin's deep intuitions) the perennial establishment of objectivity. Hence the remaining of the subject of perception identical to himself even in the presence of a creativity at play, a creativity, however, that must reveal itself as phantasmal (cf. Pier Francesco Mola, "Diana and Endymion"), i. e. as a creativity that tells of its reached invariance and for which life itself is reduced to such storytelling: i.e. to the inspection of itself as a fixed point. In this picture the recursive processes that ensure the relaxation of the imagination take, in some respects, the place of the ancient Kantian schemes.

Just as in Kant there are different types of schemes at work, also for what concerns the eigenforms we are faced with the blossoming of different types of eigenforms: boxes, frames, fractals, etc. In this respect, the real problem is the fact that Kauffman's model does not foresee the presence of specific semantic orderings at play, nor the intervention of the component related to meaning in action. From an effective point of view, at the level of cognition (and life), as we have already seen, the basic interweaving concerns the ever changing bond between reflection and simulation, i. e. between Narcissus and Marsyas. As a craftsman I must simulate possible worlds on my own skin so that the God can come to affect me. I must pose myself as a gridiron for the realization of specific concepts, only then will I be welcomed by the Muse: a creative thought will rise in me to the extent that I will come to be added. Just as thought is leavening in me, allowing me to transfigure, observation leads me to closing in a sphere, (a sphere, however, that, as shown by Pier Francesco Mola, can enclose creativity in itself, but only in a phantasmal way). I add myself as an observer and I come to be part of a Temple whose columns carry my name-symbol. Narcissus is constituted as a column and, therefore, as a replication process taking place at the level of Nature to the extent that he gives rise to the inscription of himself in the

stone and to the reading-observation of his own form as a form of distinction and visibility. It is inasmuch as the hero comes to be perceived and to perceive that he comes into existence (Berkeley) thus coming to add himself to the very heart of the world of Necessity: he presents himself both as object and symbol. He stands as an object along the giving of an eigenform. The observer-operator during his acting can only leave the form of the object as invariant. In this sense L. Kauffman can only be aware of the necessity that: "...in the course of examining the concept of reflexivity we will find that the essence of the matter is an opening into creativity."<sup>(15)</sup> This appears necessary if one wants to avoid a definitive closure in the universe of Narcissus and, therefore, in a form of creativity characterized in phantasmal terms. However, as we will see, without recourse to the simulation models it does not seem possible to outline an adequate (and not phantasmal) space for what is the process of self-organization that underlies cognitive procedures. The closure in the Narcissus world not only does not allow us to work for morphogenesis but does not even allow us to prepare what is necessary for the genesis of a renewed invariance, an invariance that cannot be solely considered as the fruit of a perennial cyclic activity (the cyclic resurrection of the hero as a flower).

At the theoretical level, the underlying goal of Louis Kauffman is to preserve, with regard to molecular Biology, the scheme relative to the central dogma as formulated by Monod (a scheme essentially linked to the first order Cybernetics) but with a view to its partial overcoming by means of the recourse to the methods proper to the second-order cybernetics. In his opinion, it is possible to achieve this overcoming making recourse to an extended application of reflexive domains able to ensure the due space for the full articulation of creativity and the continuous emergence of specific novelties. Only the success of such an attempt could lead us to the heart of a mathematically adequate theory (not only phenomenological) of that particular theoretical set-up relative to the enactive mind as proposed by Varela during his last years of life (although on largely phenomenological bases). Whilst, however, for Monod the creative evolution derives from the coupling between Chance and Necessity, for Kauffman who wants to open up to the methods of the second-order Cybernetics, the novelties must come to be born from within the system itself through an inner dialectic game such as that suggested by the functioning of that particular cellular automaton related to the Game of Life as outlined by J. Conway. However, as we will see, also an overcoming of this kind could only reveal in terms of a phantasmal "appearance" giving rise once again to a modest form of life such as that of Endymion: in fact, it will not be able to account for those bonds that weave together invariance and morphogenesis according to a real circularity, a circularity that finds its foundation first of all in the mathematics of non-standard and in that continuous passage by levels which appears inextricably linked to the arising of continuous changes with respect to the Semantics at work.

After the death of Varela another volume was published in the wake (for certain aspects) of the book published in 1999. Also at the level of this volume a central role is played by J. Petitot and his school. The volume published in 2004 (cfr. Carsetti, A. (Ed.) (2004), Seeing, Thinking and Knowing. Meaning and Self-Organization in Visual Cognition and Thought, Dordrecht, The Netherlands: Kluwer Academic Publishers) is, partially, focused on an ambitious problematic: actually, according to the Editor the volume should not be limited only to the investigation of the genesis of the embodied mind, it was also aimed to reach an initial clarification of the action exerted by the embodied meaning. This volume sees, in particular, the presence of an illuminating work by K. J. O'Regan, E. Myin and A. Noe, "Towards an Analytic Phenomenology: The Concepts of "Bodiliness" and "Grabbiness". It is precisely this article that will open new horizons through its come to determine an important change of perspective with respect to the studies concerning the genesis of the human mind. Next to this article in the book there is also a chapter by A. Carsetti dedicated to a broad analysis of the reality of the embodied meaning.

The book allows us to see, among other things, how the legacy of Varela came to spread in the context of a variety of research areas. What appears very clear, however, once we analyze this spreading in details, is that the interest of many scholars working in the wake of Varela' s heritage is increasingly converging today on the Enactivism and the Theory of the embodied mind according to the lines of the already cited paper by Varela of 1990 (whose title is already particularly illuminating in itself). It is precisely in this framework that the lines of research carried

out by O'Regan, Myin and Noe. come to open up further horizons to the gaze. We are faced with a real breakthrough but marked in conservatively. As a matter of fact, the reality of embodiment is not called in question, the setting, however, in which this reality is to be realized appears extremely diversified and complex with respect to the mathematical and epistemological tools carried out by Varela. In order to understand the ultimate nature of this intellectual shift let us remind, first of all, that starting from the years 60 of the last century we have witnessed the birth of a structuralist theory as regards the nature of numbers and, in general, of the mathematical structures underlying human cognitive activity. This birth in many respects follows the progressive affirmation of the Computational Structuralism and appears strictly linked, as we shall see, to the consequences inherent in the Tennenbaum theorem.

It dates back to 1965 the publication of the fundamental work of Benacerraf (cf. "What numbers could not be" (1965), Philosophical Review, 74: 47-73). It is precisely on the wave of the theses advocated in this paper that the Computational Structuralism has recently developed. Hence the emergence of a very interesting debate that has come to focus on the problem relative to the genesis of mind in its link with the mathematical structures, a debate that has recently seen P. Quinon and K. Zdanowski come to revisit in 2009 the main thesis by O'Regan, Myin and Noe in order to outline a new vision of what was, in Kantian terms, the main tool available to our mind, namely that "Schematism" considered by Kant as the most mysterious (and perhaps inaccessible) aspect of mind's activity. According to Kant the Schematism is the tool implemented at the level of human mind in order to guarantee the coordination of the concepts in presence of that self-organization process that identifies the genesis of the "Thinking I". The term "self-organization" is directly utilized by Kant in his major work (1781). As we have just mentioned, Kant's main intuition has been revisited by H. von Foerster at the level of the contemporary theory of self-organization, through the innovative recourse to the concept of non-trivial machines, of machines, that is, capable of self-organizing themselves.<sup>(16)</sup> An interesting example of this kind of machines is today represented, albeit at a first level of complexity, by the associative memories as investigated by Kohonen in the framework of his theory of self-organizing maps. The reference to the self-organization procedures acting at the cognitive level represents

perhaps the most appropriate key to understanding the nature of the paradigm shift that animates the work carried out by Quinon and Zdanowski in the wake of the first insights by O'Regan, Myin and Noe.

Indeed, at the level of the paper by P. Quinon and K. Zdanowski (2006) entitled: "The Intended Model of Arithmetic. An Argument from Tennenbaum Theorem" (17) we are faced with a shift that manages to field, beyond Husserl's lesson, the theses advocated by Wittgenstein in relation to his conception of meaning as use (as well as, in some respects, that theoretical perspective concerning the embodied meaning to which the last chapter of the aforesaid volume Seeing, Thinking and Knowing is dedicated). In order to understand the scope of this statement let us now to refer briefly to the lines of that scientific debate which, starting from 2004, finally came to focus on the reality of mind but with reference to the development of a Wittgensteinian perspective based on the theory of computation as well as on the mathematical instruments currently offered by the ongoing research in the field of hypercomputation. A reality this latter that appears far removed from the mathematical training by Varela. It is no longer a question of taking into account Turing's world together with the autonomous reality of a body according to the theses advocated by Varela in the aforementioned article he published in 1990, on the contrary the real issue now is to understand how from within the same reality of a computational self-organizing world that particular novelty represented by the evolution of a body together with its unique mind can come to be born. Once again, in view of being able to navigate in these mysterious waters the compass, in the first instance, can only be the one represented by the two theses by H. Atlan: a) function self-organizes together with its meaning and A. Carsetti: b) meaning self-organizes together with its creativity. That the genome cannot be identified as a simple program as claimed by Atlan, is quite clear in terms of the dynamics we can inspect at the deep biological level. As we have just said, biological information cannot be reduced to simple Shannonian information. Software identification, in turn, cannot be separated from the different phases of the self-organization process at work. At the level of the original library of programs it is necessary to identify both a depth dimension and a surface dimension; moreover we are also witnessing the continuous construction of biological (non-trivial) machines and their incessant recycling and transformation. Hence the giv-

ing of a body of Nature as hardware forged in obedience to the "programs" at play on the basis of the selection exercised by the orderings offered by the Form, a Form that will finally come to reflect itself as pure software in action. Fundamental at this level is the action of editing as connected to the processes of splicing. We have, on the one hand, genes as "predicaments" (functional modules in action) working at the DNA level and, on the other hand, pure "insights" working at the level of the Forma formans considered as a source of individuated orderings. In this sense, the different intensities (capacities) can only be revealed in precise correspondence to specific sections of the functional library in action at the biological level. In such a theoretical framework, the original "irruption" can only appear linked to a language that becomes thought in life. Conception, on the other hand, appears to be linked to that thought through forms (operating in the framework of a vision in the truth) which becomes pure language, the language of a particular body finally drowning in the waters of the Temple. Hence the possibility of a self-reflection (mirroring) on behalf of the Forma formans in the body of Nature, a mirroring which finally presides over the birth of the Lord of the garlands and the triumph of the modules characterizing the artificial life. As we have just seen, it is exactly this kind of mirroring which identifies the role played by Narcissus. Meaning can reflect itself in the body of Nature to the extent that Narcissus comes to drown in his own image in waters.

The embodied life is memory plus imagination but along the realization of that particular detachment from the original *Sylva* as performed by the Minotaur. Here we can find the roots of Narcissus' development. From the intensities to the eigenforms trough the individuation of the correct reflexive domains. Only if the original memory selforganizes as pure hardware on the basis of the intervention of the due orderings we will be faced with new imagination and new data bases at work: only in this way true morphogenesis can, then, take place.

It is Narcissus "who" allows meaning to self-conceive and the source to realize the process relative to the self-renewing "inscription". He can

\*\*\*

really offer his severed head and his achieved ability as editor only to the extent to which he recognizes himself in the inscribed design and in the reflected image. Through the coder the source assumes a reproductive capacity commensurate with a precise invariance and with the individuation of intrinsic forms which inhabit life; it inscribes itself as instructional form (information) and as a hereditary principle in action, as a source of varied complexity but compared with a hereditary apparatus which self-organizes as such in view of possible regeneration. The source which generates on the basis of self-reflection (but at the surface level) opens out, then, towards a self-reproduction process which is targeted and part of a co-evolutionary path. Whoever arrests and captures the reflection, fixing and freezing it, also makes him/herself into a reflection; the offering of him/herself as severed head to the fluxes is in sight of a new invariance and the possible emergence of ever-new specific properties. Master/mistress of the shadow, s/he guides the process of regeneration by opening up to the new "conception". The inscription and the suture of the wounds operate at the level of the becoming body; and it is by a path of perceptual activity of this kind that a world articulated in properties is finally recognized. The result is a source which, having stored a pathway (the Road as this emerges along the petrifaction), is able to code and articulate as a set of properties and recipes, thus proposing itself as mirror to itself but within the contours of incoming life. Hence the possibility of a genomic information self-mediated in the architecture of proteins.

With respect to this mainframe, if we take into consideration, for instance, visual cognition we can easily realize that vision is the end result of a construction realized in the conditions of experience. It is "direct" and organic in nature because the product of neither simple mental associations nor reversible reasoning, but, primarily, the "harmonic" and targeted articulation of specific attractors at different embedded levels. The resulting texture is experienced at the conscious level by means of self-reflection; we actually sense that it cannot be reduced to anything else, but is primary and self-constituting. We see visual objects; they have no independent existence in themselves but cannot be broken down into elementary data. Grasping information at the visual level means managing to hear, as it were, inner speech. It means first of all capturing and "playing" each time, in an inner generative language,

through progressive assimilation, selection and real metamorphosis (albeit partially and roughly) and according to "genealogical" modules, the emergent (and complex) articulation of the semantic apparatus which works at the deep level and moulds and subtends, in a mediate way, the presentation of the functional patterns at the level of the optical sieve.

What must be ensured, then, is that meaning can be extended like a thread within the file, a thread that carries the choices and the piles related to specific symmetry breakings. It is Narcissus who must donate cues in order to operate the fixing of meaning. At the end of the metamorphosis the hero will mirror himself in the motionless face of Ariadne. Now his head will be cut (cf. Caravaggio: Medusa's consciousness) and a vision will arise in accordance with the truth.

In this way, it will be possible to identify a "garland"; only on the strength of this construction can an "I" posit itself together with a sieve: a sieve in particular related to the world which is becoming visible. In this sense, the world which then comes to "dance" at the level of the eyes of my mind is impregnated with meaning. The "I" which perceives it realizes itself as the fixed point of the garland with respect to the "capturing" of the thread inside the file: i. e. inside that genealogically-modulated articulation of the file which manages to express its invariance and become "vision" (visual thinking which is also able to inspect itself), anchoring its generativity at a deep semantic dimension. The model can shape itself as such and succeed in opening the mind's eye in proportion to its ability to permit the categorial to anchor itself to (and be filled by) intuitions (which are not, however, static, but emerge as linked to a continuous process of deconstruction of the original holistic meaning). And it is exactly in relation to the adequate constitution of the channel that a sieve can effectively articulate itself and cogently realize its selective work at the informational level. This can only happen if the two selective forces, operating respectively within an ambient meaning and an ambient incompressibility, meet, and a *telos* shapes itself autonomously so as to offer itself as guide and support for the task of both capturing and "ring-threading". It is the (anchoring) rhythm-scanning of the labyrinth by the thread of meaning which allows for the opening of the eyes, and it is the truth, then, which determines and possesses them. Hence the construction of an "I" as a fixed point: the "I" of those eyes (an "I" which perceives and which exists in proportion to its ability to perceive (and "fix") according to the truth). What they see is generativity in action, its surfacing rhythm being dictated intuitively. What this also produces, however, is a file that is incarnated in a body that posits itself as "my" body, or more precisely, as the body of "my" mind: hence the progressive outlining of a meaning, "my" meaning which appears gradually pervaded by life. Concepts and schemes, intuitions and diagrams relative to orderings. Determinations of Time and diagrams of the memory. The intensities come to life, the Time comes to the truth. On the one hand schemes working on predicaments, on the other hand, orderings individuating the original insights. The categories through the schemes give rise to concepts, the Form through the orderings gives rise to intuitions. Determinations of Time and channelling of intensities. Conceptual acts and intuitive scans. The Form is filled with insights, the categorial, in turn, is populated by concepts.

Vision as emergence aims first of all to grasp (and "play") the paths and the modalities that determine the selective action, the modalities specifically relative to the revelation of the afore-mentioned semantic apparatus at the surface level but in accordance with different and successive phases of generality. These paths and modalities thus manage to "speak" through my own fibers. It is exactly through a similar self-organizing process, characterized by the presence of a double-selection mechanism, that our mind can partially manage to perceive (and assimilate) depth information in an objective way. Here we can find the ultimate roots of its genesis. The extent to which the network-model succeeds, albeit partially, in encapsulating the secret cipher of this articulation through a specific chain of functional modules determines the model's ability to see with the mind's eye as well as, in perspective, the successive irruption of new patterns of creativity. Only if the Minotaur manages to open his eyes, can Marsyas successively perform his experimentum crucis. To assimilate and see, the system must first "think" internally of its secret abstract "capacities", and then posit itself as a channel (through the precise indication of forms of potential coagulum) for the process of opening and anchoring of depth information. This process then realizes itself gradually into the system's fibers, via possible selection, in accordance with the coagulum possibilities and the meaningful connections offered successively by the system itself (as immersed in its meaning).

The revelation and channelling procedures thus emerge as an essential and integrant part of a larger and coupled process of self-organization. In connection with this process we can ascertain the successive edification of an I-subject conceived as a progressively wrought work of abstraction, unification, and emergence. The fixed points which manage to articulate themselves within this channel, at the level of the trajectories of neural dynamics, represent the real bases on which the "I" can graft and progressively constitute itself. The I-subject (the observer) can thus perceive to the extent in which the single visual perceptions are the end result of a coupled process which, through selection, finally leads the original Source to articulate and present itself as true invariance and as "harmony" within (and through) the architectures of reflection, imagination, computation and vision, at the level of the effective constitution of a body and "its" intelligence: the body of "my" mind. These perceptions are (partially) veridical, direct, and irreducible. They exist not only in themselves, but, on the contrary, also for the "I", but simultaneously constitute the primary departure-point for every successive form of reasoning perpetrated by the agent. As an observer I shall thus witness Natura naturata since I have connected functional forms at the semantic level according to a successful and coherent "score". In this sense at the level of the reflexivity proper to the system the eigenforms reveal themselves as an integrant part of that self-organization process which constitutes the real engine of visual cognition, a process that the *telos* itself can manage to "imagine" only along the progressive unfolding of its reflexive tools. Without the individuation of the "I" and the composition of the eigenforms no perceptual activity is really possible.<sup>(18)</sup>

The multiple unfolding of the eigenforms will be tailored to the symmetry breakings that manage to be realized at the level of meaning. They come to constitute themselves as fixed points in the process of construction of the structures of the operator. In this sense, they present themselves as the real bases of my own perceptual operations and, therefore, "preside" at the identification of the objects in the world. The forms arise from the determinations of the embodiment taking place by means of the "infixions" offered by Ariadne. When the garland closes and embraces the Minotaur that embodies and reflects himself, we have the emergence of vision. Narcissus-Minotaur will finally be able to recognize himself as an invariant reality and a source of replication through

96

his rendering to the "stone", i.e his realization as living being in the world and his becoming an integrant part of the ruler (through self-adjunction).

In accordance with these intuitions, we may tentatively consider, from the more general point of view of contemporary Self-organization theory, the network of meaningful (and "intelligent") causal "programs" living at the level of our body as a complex one which forms, articulates, and develops, functionally, within a "coupled universe" characterized by the existence of two interacting selective forces. This network gradually posits itself as the real instrument for the actual emergence of meaning and the simultaneous, if indirect, surfacing of an "acting I": as the basic instrument, in other words, for the perception of real and meaningful processes, of "objects" possessing meaning, aims, intentions, etc.: above all, of objects possessing an inner plan and linked to the progressive expression of a specific cognitive action.

As we have just said, the mechanism which "extracts" pure intuitions from the underlying formal co-ordination activity, if parallel to the development of the *telos* as editor with respect to the coder, is necessarily linked to the continuous emergence of new mathematical moves at the level of the neural system's cognitive elaboration, This consideration inviting the revisiting of a number of Kantian hypotheses. It would appear, for instance, to be necessary not only to reread Kant in an evolutionistic key (cf. e.g., K. Lorenz), but also with reference to other speculative themes like, for instance, the indissoluble link existing between life and cognition and between chance and necessity. Taking into consideration coder's action opens up a new and different relationship with the processes of mathematical invention, making it necessary, for example, to explore second-order territories, the very realm of non-standard mathematics as well as the dialectics between observer and observed reality.<sup>(19)</sup>

*Pace* Kant, at the level of a biological cognitive system sensibility is not a simple interface between absolute chance and an invariant intellectual order. On the contrary, the reference procedures, if successful, are able to modulate canalization and create the basis for the appearance of ever-new frames of incompressibility through morphogenesis. This is not a question of discovering and directly exploring (according, for in-

stance, to Putnam's conception) new "territories", but of offering ourselves as the matrix and arch through which they can spring autonomously in accordance with ever increasing levels of complexity. There is no casual autonomous process already in existence, and no possible selection and synthesis activity *via* a possible "remnant" through reference procedures considered as a form of simple regimentation. These procedures are in actual fact functional to the construction and irruption of new incompressibility: meaning, as *Forma formans*, offers the possibility of creating a holistic anchorage, and is exactly what allows the categorial apparatus to emerge and act according to a coherent "arborization". From the encounter of Ariadne with Narcissus we shall have the flowering of forms, the possibility to perceive by fixed points, the birth of specific structures at the level of the operator.

The new invention, which is born then shapes and opens the (new) eyes of mind: I see as a mind because new meaning is able to articulate and take root through me. As J. Petitot correctly remarks, according to Kant the pure intuitions are : « "abstraites de l'action même par laquelle l'esprit coordonne, selon des lois permanentes, ses sensations" (Dissertation, 177). Or, cette coordination est elle-même innée et fonctionne comme un fondement de l'acquisition»<sup>(20)</sup>. In this sense, the space appears as a format, the very basis of spatial intuition is innate. According to Kant, it is a condition of a subject knowing anything that the things it knows should be unified in a single consciousness. Kant calls this consciousness the transcendental unity of apperception. Kant writes that this unity comes about "not simply through my accompanying each representation with consciousness, but only in so far as I conjoin one representation with another." (B 133, p. 153 in Kemp Smith, A Commentary to Kant's 'Critique of Pure Reason' (London: Macmillan, 1918) ). In this sense, all coherent consciousness, hence all knowledge of anything, presupposes not just an original unity, but original, conceptual acts of possible combination to produce such unity. This means that some concepts are a priori. They cannot possibly have been derived from experience, since without them there would have been no original unity of experience. Just as Kant identifies in this way the existence of a priori conceptual acts (by coniunctio) living at the level of the potential intellect, he also identifies, as we have just said, the existence of innate tessellations (by orderings) at the level of the Form. It is in dependence of the determinations operated by the schemes on the Form of Time (in connection with the operations of categories-capacities) that we will have the intuitions that populate our perception of the world

The schemes allow a space-time tessellation according to intuitions. Here are the processes that lead to the articulation of specific eigenforms. From here, moreover, the profiling of that digital image in which Narcissus drowns reaching its objectivity. Starting from the capacities and the action of Grace we reach the eigenforms and the invariance. This, however, allows the new conception as well as the path pursued by the Work, the same come to burn in the air, in the close exploration of its most hidden secrets, by the Painter (as De Nittis states). Here is the current extroversion and the related eigenvalues. Hence the possible trigger and the new incarnation related to it (an incarnation that arises from irruption and not from conception). Capacities as intensities and as articulations of thought in life that are made, therefore, to eigenforms. On the opposite side, however, we are faced with meanings that are made to eigenvalues. Meaning as a trigger for the incarnation following the pyre, creativity as a condition for the abstraction following the petrifaction.

In this sense, at the biological level, as we have just seen, what is innate is the result of an evolutive process and is "programmed" by natural selection. Natural selection is the coder (once linked to the emergence of meaning): at this level the emergence process is indissolubly correlated to the continuous construction of new formats in accordance with the unfolding of ever new mathematics, a mathematics that necessarily moulds coder's activity. Hence the necessity of articulating and inventing a mathematics capable of engraving itself in an evolutive landscape in accordance with the opening up of meaning. As we have just said and as we shall see in the following chapters, the realms of non standardmodels and non-standard analysis represent today a fruitful perspective in order to point out, in mathematical terms, some of the basic concepts concerning the articulation of an adequate intentional information theory. This individuation, on the other side, presents itself not only as an important theoretical achievement but also as one of the essential bases of our very evolution as intelligent organisms. Here we can find the basis for the development of the mathematical language related to a new science: Metabiology.

#### Notes

- (1) Cf. Autopoiesi e teoria dei sistemi viventi, (1984) La Nuova Critica, 64.
- (2) Cf. Spencer Brown, G. (1969), Laws of Form, London, Allen and Unwin.
- (3) Cf. Kauffman L. H. (2018), "Mathematical themes of Francisco Varela", La Nuova Critica, 65-66, p. 72.
- (4) Cf. Fuchs C.H. (2016), "On Participatory Realism", <u>arXiv:1601.04360v3</u> [quant-ph].
- (5) Cf. Varela F., Thompson E. and Rosch E. (1991), New York, MIT, p. 8.
- (6) Cf. Varela F., Thompson E. and Rosch E. (1991), New York, MIT, p. 9.
- Cf. Merleau-Ponty M., (1958), *Phenomenology of Perception*, New York, Routledge, x-xi.
- (8) Cf. Dennett D.C. (1993), "Review of F. Varela, E. Thompson and E. Rosch, The Embodied Mind, American Journal of Psychology, 106, p 121.
- (9) Cf. Dennett D.C. (1993), "Review of F. Varela, E. Thompson and E. Rosch, *The Embodied Mind*," (1993) *American Journal of Psychology*, 106, p 121.
- (10) Cf. Varela F. (1995), "The Emergent Self" in John Brockman, *The Third Culture*, New York, Simon & Schuster, p. 154.
- (11) Cf. Varela F. (1995), "The Emergent Self" in John Brockman, *The Third Culture*, New York, Simon & Schuster, p. 155.
- (12) Cf. Kauffman L. (2009), "Reflexivity and Eigenforms", *Constructivist Foundations*, 4, 3, p. 123.
- (13) Cf. Kauffman L., "Eigenform and Reflexivity" (2016) *Constructivist Foundations*, 12, 3, p. 250.
- (14) Cf. Kauffman L. (2009), "Reflexivity and Eigenforms", *Constructivist Foundations*, 4, 3, p. 127.
- (15) Cf. Kauffman L.(2009), "Reflexivity and Eigenforms" (2009), Constructivist Foundations, 4, 3, p. 121.
- (16) Cf. von Foerster H. (1981) "On constructing a reality" in *Observing Systems*, Intersystems Publications, pp. 288-309; Kohonen T., (1995) *Self-Organizing Maps*, Springer, New York, 1995.
- (17) Cf. Quinon P. and Zdanowski K. (2006), The Intended Model of Arithmetic. An Argument from Tennenbaum's Theorem, *http://www.impan.pl/\_kz/\_les/PQKZ Tenn.pdf*.

- (18) Cf. Carsetti A. (1987), "Teoria algoritmica dell'informazione e sistemi biologici", La Nuova Critica, 3-4:37-66; Carsetti A. (2000), "Randomness, Information and Meaningful Complexity: Some Remarks About the Emergence of Biological Structures", La Nuova Critica, 36: 47-109.
- (19) Cf. Carsetti A. (2013) *Epistemic complexity and knowledge construction*. New York, Springer.
- (20) Cf. Petitot J. (2008), *Neurogéometrie de la vision*", Paris, Ecole Polytechnique, p. 397.

## HUMBERTO R. MATURANA

# L'ILLUSIONE DELLA PERCEZIONE, OVVERO LA CHIUSURA OPERATIVA DEL SISTEMA NERVOSO

Se prestiamo attenzione alle nostre esperienze quotidiane, non possiamo fare a meno di notare che un oggetto che intercetti simultaneamente due luci, una rossa e una bianca, proietta due ombre, una delle quali appare rossa e l'altra verde. Se ci chiediamo come accada che una delle due ombre appaia verde, dal momento che non c'è nessuna luce verde, la risposta di solito è che si tratta di un'illusione, il verde è un colore indotto dal contrasto simultaneo. Questo genere di risposta sarebbe accettabile solo nel caso in cui le cosiddette illusioni non fossero esperienze percettive altrettanto potenti quanto lo sono le percezioni propriamente dette, ed esse lo sono. Infatti, possiamo affermare che una particolare esperienza sensoriale è un'illusione e non una percezione solo facendo affidamento, in quanto percezione, su un'altra esperienza sensoriale al posto di quella. Nel caso in questione di ombre colorate, consideriamo il verde come un'illusione solo se accettiamo come verdetto l'esperienza sensoriale che conseguiamo mediante uno spettroscopio, secondo cui riceviamo esclusivamente luce bianca dall'area in cui vediamo verde. Ma, se ammettiamo che certe esperienze sensoriali siano illusioni, come possiamo sapere o ammettere che altre non lo siano? La risposta corrente a tale questione è che in un modo o nell'altro il sistema nervoso ottiene dall'ambiente i dati o le informazioni necessarie a consentirgli di elaborare una rappresentazione del mondo esterno che gli permetta di convalidare la distinzione tra percezione e illusione.

Io la penso diversamente e la mia risposta è che percezione e illusione non sono distinguibili nel funzionamento del sistema nervoso, e che qualsiasi distinzione tra di esse può nascere esclusivamente nel dominio delle descrizioni. Ciò significa che, secondo me, il sistema nervoso non funziona come un organo che usa le informazioni sensoriali per costruire una rappresentazione dell'ambiente, che poi utilizza per calcolare il comportamento dell'organismo. Lasciatemi spiegare cosa intendo rispondendo alle seguenti domande:

A. In che modo il sistema nervoso funziona in quanto sistema?

B. In che modo il sistema nervoso prende parte alla generazione di una condotta adeguata dell'organismo?

# A. In che modo il sistema nervoso funziona in quanto sistema?

Il sistema nervoso: In quanto componente dell'organismo, il sistema nervoso è un sistema a struttura determinante. Vale a dire che, qualunque cosa accada al suo interno, tanto in seguito alle sue dinamiche interne quanto alla sua partecipazione alle dinamiche e alle interazioni interne all'organismo che esso integra, esso è determinato dal funzionamento delle sue componenti nell'azione reciproca delle loro proprietà, come avviene in ogni sistema meccanicistico. Che le cose stiano così è messo in luce dalle osservazioni neurofisiologiche, cliniche e sperimentali, che mostrano in che modo i suoi componenti funzionino. Pertanto, la differenza tra il sistema nervoso e altri sistemi a struttura determinante come un telaio, un'automobile o un computer si basa sulla sua organizzazione e sulla natura dei suoi componenti. In altre parole, il sistema nervoso differisce da altri sistemi meccanistici per le relazioni esistenti tra i componenti che ne fanno un sistema nervoso e che devono restare invarianti perché esso rimanga un sistema nervoso nel corso delle sue continue trasformazioni, nonché per i tipi di componenti che lo costituiscono. Lasciatemi chiarire questo aspetto.

Ogni sistema a struttura determinante ha un'organizzazione e una struttura. L'organizzazione di un sistema a struttura determinante è data dalla configurazione delle relazioni esistenti tra i suoi componenti che ne fanno un sistema di un particolare tipo, vale a dire un'automobile, una fabbrica o un sistema nervoso. L'identità di classe di un sistema si basa sulla sua organizzazione; se cambia l'organizzazione di un sistema, allora cambia anche la sua identità di classe, cioè esso si disintegra in quanto sistema di un certo tipo e al suo posto appare qualcosa d'altro. La struttura di un sistema a struttura determinante è la sua costituzione effettiva in quanto entità particolare, e comprende i suoi componenti effettivi e le relazioni concrete che essi stabiliscono, incluse quelle relative all'organizzazione. Pertanto, dal momento che la struttura di un sistema a struttura determinante comprende più relazioni di quelle dell'organizzazione, la sua struttura può cambiare senza che esso si disintegri fintantoché non vengano cambiate le relazioni della sua organizzazione. Se la struttura di un sistema a struttura determinante cambia senza che esso perda la sua identità di classe, cambiano le sue proprietà in quanto sistema particolare, ma rimangono invarianti le sue proprietà e caratteristiche in quanto sistema appartenente ad una determinata classe. In queste circostanze accade che ogni sistema a struttura determinante abbia:

- *a)* un dominio di cambiamenti di stato in quanto dominio di cambiamenti di struttura senza perdita dell'identità di classe;
- *b)* un dominio di cambiamenti distruttivi in quanto dominio di cambiamenti di struttura con perdita dell'identità di classe;
- c) un dominio di perturbazioni in quanto dominio di possibili interazioni strutturali che innescano in esso un cambiamento di stato, e
- *d*) un dominio di disintegrazioni in quanto dominio di possibili interazioni strutturali che innescano in esso un cambiamento distruttivo.

Accade che in un sistema a struttura determinante, ogni cambiamento strutturale che si possa prendere in considerazione si verifichi in seguito a innesco prodotto dalle interazioni dei suoi componenti, sia internamente l'uno con l'altro, che esternamente con i componenti del medium in cui esso funziona come unità. In altre parole, la struttura di un sistema a struttura determinante determina sia il corso dei suoi cambiamenti strutturali (casi a e b citati sopra), sia quali configurazioni strutturali del medium possono interagire con esso e innescare in esso un cambiamento strutturale (casi c e d citati sopra).

Da tutto ciò consegue che qualsiasi spiegazione del funzionamento del sistema nervoso deve rispettare le limitazioni fondamentali in base alle quali esso deve funzionare come un sistema dinamico a struttura determinante, e in particolare: 1) che ogni cambiamento strutturale che si verifichi in esso cambierà o i suoi domini di cambiamenti strutturali ae b, o quelli delle interazioni c e d, oppure entrambi; 2) che tutti i suoi cambiamenti strutturali sorgeranno in quanto innescati dalle interazioni dei suoi componenti, o in seguito alla sua stessa dinamica, o attraverso la sua partecipazione alle interazioni dell'organismo; e 3) che la sequenza effettiva delle interazioni tra i suoi componenti selezionerà in esso, mediante innescamenti successivi, una particolare linea di cambiamenti strutturali in quanto sua epigenesi.

Organizzazione e struttura del sistema nervoso: Se il sistema nervoso è un sistema a struttura determinante, in che modo partecipa alla produzione di un comportamento adeguato da parte dell'organismo in un medium che cambia? Io sostengo che è possibile rispondere in modo corretto a questa domanda solo se si ammette che: 1) il sistema nervoso è organizzato come un reticolo chiuso di elementi neurali in interazione, tali che attraverso le loro mutevoli relazioni di attività si generano soltanto ulteriori cambiamenti delle relazioni di attività tra loro stessi; 2) esso è continuamente sottoposto a cambiamenti strutturali che seguono una sequenza selezionata dalle interazioni interne ed esterne dei suoi componenti che li innescano. Per essere chiaro, io propongo che sia una caratteristica dell'organizzazione del sistema nervoso in quanto componente dell'organismo che esso funzioni come un reticolo chiuso di elementi in interazione in modo tale che ciascun cambiamento dell'attività che si verifichi in esso conduca sempre ad ulteriori cambiamenti dell'attività in esso. Io propongo che il sistema nervoso nel funzionare come un sistema dinamico di relazioni mutevoli di attività non abbia *input* o *output*, ma che esso generi esclusivamente gli schemi mutevoli (ricorrenti o meno)

106

delle relazioni interne che la sua struttura mutevole determina ad ogni istante. Inoltre, io propongo che la struttura del sistema nervoso in quanto componente dell'organismo sia quella di un reticolo di effettori, di recettori e di neuroni (che io chiamo tutti elementi neuronali) che si connettono reciprocamente in molti modi differenti, specifici di ciascuna specie, e che questo reticolo si chiuda in sé stesso anche attraverso il medium in una sorta di sinapsi sensoriale effettoria. In altre parole, propongo che noi, in quanto osservatori, ci collochiamo nell' interruzione sinaptica sensoriale effettoria dell'organismo, e che ciò che per noi rappresenta un medium altamente strutturato sia per il funzionamento del sistema nervoso soltanto una linea informe di chiusura come ogni altra interruzione sinaptica.

Da ciò che ho detto consegue che il medium non entra come tale nel funzionamento del sistema nervoso. Ciò che è peculiare del medium è che noi in quanto osservatori ci collochiamo in esso e che le strutture che consideriamo come ambiente sono importanti per il nostro funzionamento come osservatori, ma non lo sono per il funzionamento del nostro sistema nervoso in quanto reticolo neuronale che genera soltanto cambiamenti delle relazioni di attività tra i suoi componenti. Secondo questa concezione, le superfici effettorie e sensoriali che vediamo nell'organismo sono punti in cui la struttura del suo sistema nervoso è aperta a subire interazioni strutturali nel nostro dominio di osservazione (che chiamiamo interazioni con il medium), ma in cui esso rimane chiuso dal punto di vista operativo. In altre parole, ogni interazione dell'organismo al livello di una superficie sensoriale innesca un cambiamento strutturale nel suo sistema nervoso, che produce un cambiamento nella dinamica degli stati di questo sistema perché è cambiata la sua struttura e non perché esso abbia ricevuto qualcosa come un input informazionale dal medium. Pertanto, per il funzionamento del sistema nervoso è irrilevante se questo cambiamento strutturale sia innescato da una qualche caratteristica strutturale che noi vediamo nel medium o da un cambiamento strutturale che si verifica nelle superfici effettorie dell'organismo. Tale distinzione si realizza nel dominio delle nostre osservazioni e non nel dominio di funzionamento del sistema nervoso. Tuttavia, benché la struttura del medium sia irrilevante per il funzionamento del sistema nervoso in quanto reticolo chiuso di mutevoli relazioni di attività, essa non lo è per la dinamica del suo cambiamento strutturale.

## Humberto R. Maturana

Accoppiamento strutturale: Il funzionamento del sistema nervoso consiste nella generazione continua di schemi mutevoli di relazioni di attività tra i suoi componenti, e si realizza attraverso i cambiamenti strutturali che questi innescano reciprocamente mediante le loro interazioni, in un modo che segue un corso determinato, ad ogni istante, dalla sua struttura a reticolo. Queste relazioni mutevoli di attività tra gli elementi neuronali del sistema nervoso, tuttavia, possono implicare altri cambiamenti strutturali in essi, oltre a quelli propri della loro partecipazione al suo funzionamento. Per essere chiari su questo aspetto, quando un neurofisiologo studia il sistema nervoso egli osserva quattro tipi fondamentali di fenomeni che può descrivere o meno in termini strutturali, ma che provengono tutti da cambiamenti strutturali degli elementi neuronali:

a) Cambiamenti dei potenziali di membrana degli elementi neuronali che, in diminuzione o propagati, innescano nelle stesse cellule processi locali aggiuntivi che dànno luogo all'innescamento di cambiamenti nei potenziali di membrana di altri elementi neuronali in luoghi determinati dal modo in cui questi ultimi si connettono reciprocamente tra di loro. I processi locali citati possono consistere in cambiamenti presinaptici, che eventualmente innescano cambiamenti post-sinaptici, cambiamenti di forma degli elementi neuronali stessi che producono il movimento di una parte dell'organismo e il conseguente innescamento di ulteriori cambiamenti di stato in alcune delle sue aree di ricezione, oppure cambiamenti nella secrezione di sostanze che, alla fine, producono anch'essi l'innescamento di un cambiamento strutturale in una certa superficie recettoria. Ne risulta che il sistema nervoso si trova in uno stato di cambiamento strutturale continuo e ricorsivo e che questo si produce durante e attraverso il suo funzionamento in quanto reticolo neuronale che genera schemi stabili o mutevoli di relazioni di attività determinati in ogni momento dalla sua struttura effettiva a rete chiusa. Un neurofisiologo si riferirebbe a questi cambiamenti strutturali come a cambiamenti strutturali propri del funzionamento del sistema nervoso in quanto rete di relazioni mutevoli di attività.

b) Cambiamenti di struttura del sistema nervoso mediante cambiamenti di struttura dei suoi componenti, innescati da sostanze liberate
localmente dagli elementi neuronali stessi conformemente ai processi di interazione tra di essi che non sono mediati dai cambiamenti dei potenziali di membrana a cui si è fatto riferimento in *a*, ma che sono distribuiti a seconda della collocazione dei loro contatti in quanto componenti del reticolo neuronale. Le modulazioni strutturali reciproche, risultanti da questi innescamenti scambievoli di cambiamenti strutturali tra gli elementi del reticolo strutturale, vengono chiamate effetti trofici. Esse, comunque, non si verificano in quanto parte del funzionamento effettivo del sistema nervoso come reticolo neuronale, come nel caso dei cambiamenti strutturali citati nel paragrafo precedente *a*, e possono essere meglio considerate in quanto realizzantesi attraverso processi strutturali ortogonali rispetto a questo funzionamento.

c) Cambiamenti di struttura del sistema nervoso innescati da sostanze liberate nell'organismo da cellule che non appartengono propriamente al reticolo neuronale, e che pertanto possono trovarsi lontane dalle cellule che reagiscono ad esse. Queste sostanze che vengono distribuite a tutto l'organismo attraverso il sangue e la linfa, innescano cambiamenti strutturali in molti tipi diversi di cellule oltre a quelle del sistema nervoso. Di solito vengono chiamate ormoni e i cambiamenti strutturali che esse innescano vengono generalmente considerati come processi di differenziazione cellulare. È evidente che nel sistema nervoso questi cambiamenti strutturali si realizzano attraverso processi ortogonali rispetto al suo funzionamento vero e proprio come reticolo neuronale.

d) Cambiamenti di struttura del sistema nervoso innescati da sostanze o agenti fisici che non sono prodotti nell'organismo e che appartengono al medium, oppure prodotti nell'organismo ma operativamente esterni al sistema nervoso e che interagiscono con esso in quanto parte del suo medium. Ovviamente, questi cambiamenti strutturali sono anche causati da processi strutturali ortogonali rispetto al funzionamento del sistema nervoso in quanto rete neuronale chiusa.

Esistono, pertanto, due tipi fondamentali di cambiamenti strutturali a cui il sistema nervoso è sottoposto: quelli implicati nel suo funzionamento in quanto rete neuronale e quelli implicati nella sua dinamica in quanto sistema cellulare. Il primo tipo si produce attraverso la generazione di relazioni di attività neuronali nel funzionamento effettivo del siste-

ma nervoso in quanto rete neuronale chiusa; il secondo attraverso processi che implicano gli elementi neuronali in modo ortogonale rispetto a questo funzionamento. La separazione tra questi due tipi di cambiamenti strutturali è operativa e non strutturale, benché le conseguenze siano strutturali. Vediamole. Il primo tipo di cambiamenti strutturali può essere ciclico, cosicché le strutture dei componenti del sistema nervoso ritornano periodicamente ad un certo stato iniziale dopo una serie di cambiamenti; tuttavia, possono anche non essere ciclici. Quando non lo sono, conducono di fatto a cambiamenti del secondo tipo attraverso situazioni del tipo d precedentemente menzionate. Non essendo reversibili, i cambiamenti strutturali di questo secondo genere producono una trasformazione strutturale continua del sistema nervoso, che segue un corso ricorrentemente selezionato dai cambiamenti di stato e dalle interazioni dell'organismo che esso integra. Di conseguenza, questo processo mantiene il sistema nervoso in una continua congruenza strutturale dinamica con l'organismo in quanto entità dinamica in un medium, oppure lo conduce alla disintegrazione. Ne consegue anche che questo processo mantiene il sistema nervoso in una dinamica di continue trasformazioni reversibili e irreversibili del suo dominio operativo, che è necessariamente congruente con il dominio operativo dell'organismo, oppure lo conduce alla disintegrazione.

In generale, io chiamo accoppiamento strutturale la congruenza strutturale che, sia come processo dinamico che come condizione della complementarietà statica o dinamica, si produce necessariamente quando due sistemi interagiscono ricorrentemente l'un l'altro, selezionando reciprocamente i loro rispettivi cambiamenti strutturali, e che persiste fintantoché essi interagiscono senza disintegrarsi. Ritengo inoltre che è attraverso l'accoppiamento strutturale che il sistema nervoso e l'organismo che esso integra funzionano necessariamente in congruenza dinamica nel corso della storia di integrazioni di quest'ultimo.

## B. In che modo il sistema nervoso prende parte alla generazione di una condotta adeguata dell'organismo?

I cambiamenti di stato dell'organismo in un ambiente appaiono ad un osservatore sotto forma di condotta. Se l'osservatore adotta come riferimento per le sue osservazioni la sopravvivenza dell'organismo in quanto tale, allora egli può vedere che, poiché l'organismo insieme con il suo sistema nervoso può sottostare solo a cambiamenti strutturali determinati dalla sua struttura, la sua condotta è adeguata soltanto se la sua struttura dinamica rimane congruente con la struttura dinamica del suo medium, e può continuare ad interagire con esso senza disintegrarsi. Ovvero, la condotta dell'organismo in quanto tale è adeguata solo fintantoché esso (incluso il suo sistema nervoso) rimane in accoppiamento strutturale con il medium. Allora, la riposta alla domanda posta sopra è che il sistema nervoso prende parte alla generazione di una condotta adeguata dell'organismo, specificando molti dei suoi cambiamenti di stato durante le sue interazioni, in quanto è in accoppiamento strutturale con esso e, attraverso di esso, con il suo medium. In altre parole, la condotta adeguata di un particolare organismo è il risultato della storia particolare di interazioni selettive ininterrotte tra un organismo dotato di sistema nervoso e il suo medium, che si è originata nel corso della filogenesi dell'organismo in questione, e viene portata a compimento al presente nella sua ontogenesi.

*a)* Dal momento che il funzionamento di un sistema a struttura determinante è determinato dalla sua struttura, in esso non vengono compiuti errori. Esso fa ciò che fa e non può fare altrimenti. Così, in senso stretto, per l'organismo (incluso il suo sistema nervoso) non esiste nessuna condotta adeguata o inadeguata, esso può solo sottostare a cambiamenti strutturali fintantoché conserva la sua identità. È per l'osservatore che guarda l'organismo in un contesto da lui definito, che la condotta dell'organismo soddisfa o meno le sue aspettive ed è o non è adeguata. È solo per lui, e può essere soio per lui, che l'organismo compie un errore. Quando il dominio delle aspettative e quello dell'accoppiamento strutturale in cui l'organismo viene osservatore adotta la prospettiva del medium per valutare la sua condotta, allora l'espressione

di una condotta adeguata è la sopravvivenza. Tuttavia, l'osservatore può scegliere qualsiasi altro riferimento per giudicare ciò che egli considera come condotta dell'organismo, e così facendo dare per scontata la sua sopravvivenza e vedere un errore o una condotta adeguata nell'altro dominio di sua scelta,

b) Per quanto riguarda il funzionamento del sistema nervoso non esistono fenomeni interni o esterni; tutto avviene in esso. È una caratteristica costitutiva del sistema nervoso che esso non possa distinguere attraverso il suo funzionamento tra percezione e ifiusione. Una tale distinzione può essere compiuta solo da un osservatore che sia esterno all'organismo. La congruenza delle distinzioni tra percezioni e illusioni che si può riscontrare in una comunità è il risultato dell'accoppiamento strutturale reciproco dei membri della comunità.

c) Un osservatore può parlare della complessità del dominio degli stati del sistema nervoso e della complessità del dominio delle condotte dell'organismo. La complessità del primo dominio è determinata dalla struttura del sistema nervoso in quanto rete chiusa, e in linea di principio è la stessa per tutti gli organismi della stessa specie. La complessità del secondo dominio dipende dalla storia di accoppiamento strutturale dell'organismo ed è contestuale e storica, cosicché, in linea di principio, varia per i differenti membri della stessa specie. Nell'uomo la maggior parte di questa complessità è sociale.

d) Dal momento che il medium di un sistema è determinato in quanto suo dominio di interazioni dalla sua struttura, e dal momento che due sistemi diventano strutturalmente accoppiati mediante le loro interazioni, un organismo sarà sottoposto ad accoppiamento strutturale in tutte le dimensioni delle interazioni che esso può permettersi attraverso le proprietà dei suoi componenti e attraverso le circostanze delle sue interazioni. Pertanto, un osservatore vedrà che un organismo partecipa, attraverso il suo accoppiamento strutturale in un dominio sociale (in quanto dominio di accoppiamenti strutturali ontogenetici tra organismi), alle condotte sociali che egli può descrivere come condotte in un dominio di astrazioni, tenuto conto delle circostanze storiche della loro espletazione.

e) Il sistema nervoso non crea la condotta dell'organismo, tuttavia, aumentando la diversità dei possibili stati interni dell'organismo, esso ac-

cresce il dominio delle condotte possibili da parte dell'organismo. E, inoltre, aumentando il dominio delle condotte possibili da parte dell'organismo, il sistema nervoso accresce la diversità dei possibili accoppiamenti strutturali dell'organismo nel dominio sociale.

Nessun organismo si trova dove si trova per caso, e nessun organismo (incluso il sistema nervoso) ha per caso la struttura che ha. La condotta adeguata in un certo dominio è, in ogni momento, il risultato dell'accoppiamento strutturale dell'organismo in quel dominio, e non la creazione del sistema nervoso. Il sistema nervoso attraverso la sua struttura plastica e la chiusura operativa fornisce all'organismo un dominio enorme e sempre mutevole di stati interni e rende in teoria possibile che esso rimanga in accoppiamento strutturale in un medium mutevole multidimensionale. La situazione attuale di un organismo, per quanto complessa possa apparire la sua condotta, è sempre il risultato della storia filogenetica e ontogenetica di accoppiamento strutturale a cui esso appartiene.

(Traduzione di Barbara Continenza)