THE ROYAL SOCIETY

J. R. Soc. Interface (2004) **1**, 99–107
DOI: 10.1098/rsif.2004.0012

Autopoiesis with or without cognition: defining life at its edge

Michel Bitbol[1,†] and Pier Luigi Luisi[2]

[1] *CREA, CNRS/Ecole Polytechnique, 1, rue Descartes, 75005 Paris, France*
[2] *Department of Biology, University of Roma 3, Rome, Italy*

This paper examines two questions related to autopoiesis as a theory for minimal life: (i) the relation between autopoiesis and cognition; and (ii) the question as to whether autopoiesis is the necessary and sufficient condition for life. First, we consider the concept of cognition in the spirit of Maturana and Varela: in contradistinction to the representationalistic point of view, cognition is construed as interaction between and mutual definition of a living unit and its environment. The most direct form of cognition for a cell is thus metabolism itself, which necessarily implies exchange with the environment and therefore a simultaneous coming to being for the organism and for the environment. A second level of cognition is recognized in the adaptation of the living unit to new foreign molecules, by way of a change in its metabolic pattern. We draw here an analogy with the ideas developed by Piaget, who recognizes in cognition the two distinct steps of assimilation and accommodation. While *assimilation* is the equivalent of uptake and exchange of usual metabolites, *accommodation* corresponds to biological adaptation, which in turn is the basis for evolution. By comparing a micro-organism with a vesicle that uptakes a precursor for its own self-reproduction, we arrive at the conclusion that (a) the very lowest level of cognition is the condition for life, and (b) the lowest level of cognition does not reduce to the lowest level of autopoiesis. As a consequence, autopoiesis alone is only a necessary, but not sufficient, condition for life. The broader consequences of this analysis of cognition for minimal living systems are considered.

Keywords: artificial life; self-organization; metabolism; cognition; lipidic vesicles

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).

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

† Author for correspondence (michel.bitbol@wanadoo.fr).

Received 14 June 2004
Accepted 24 September 2004
Published online 20 October 2004

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.

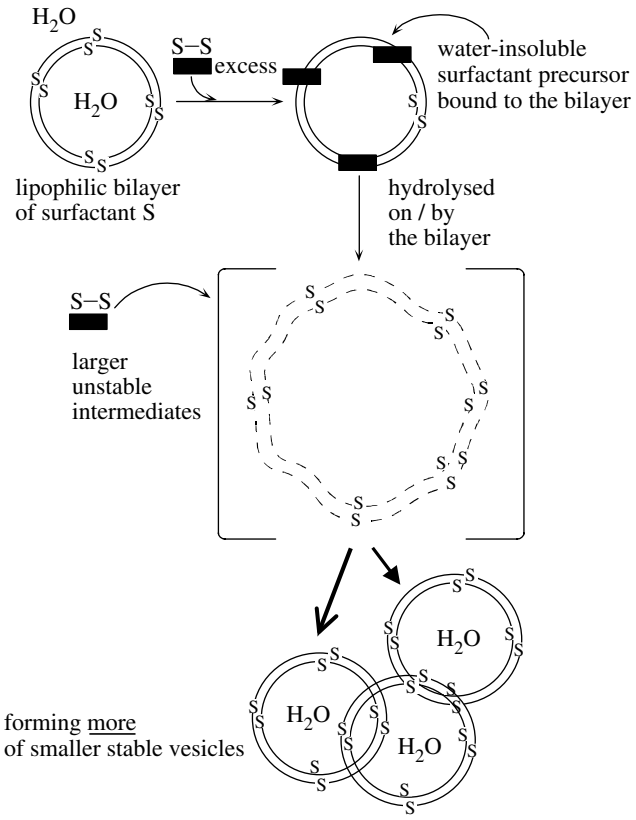

Figure 1. The autopoietic self-reproduction of vesicles.

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 hydrolysed 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 three-dimensional tessellation automaton.

While the system in figure 1 corresponds to autopoiesis in a self-reproducing 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.[1] From this remark (and irrespective of the persistently dualist undertones of its statement) one is led to contest the

[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.

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 full-blown 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 primarily 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[2] 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

[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 1989*b*). On the concept of 'structural coupling', the main sources are Maturana & Varela (1980, 1986).

the domain of biology and cognition.[3] 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.

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, non-nutrient 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 self-protecting 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 defining the target of selection can be found in a recent review by McMullin (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.

[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).

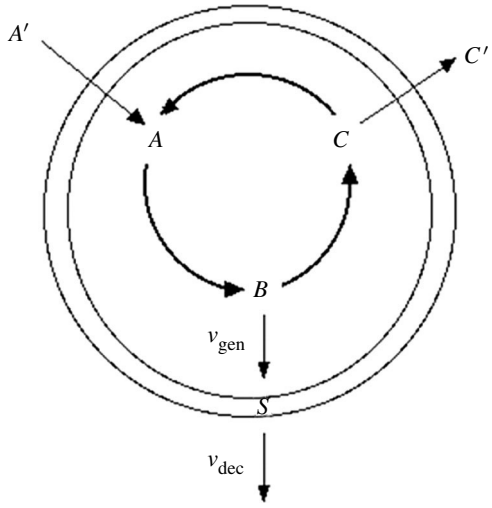

Figure 2. The membrane (S) is formed from B through a process characterized by a velocity v_{gen}. Then, S decays with velocity v_{dec}. The precursor metabolite A' enters from the environment. Furthermore, C decays into C' that is eventually expelled.

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' 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).[4] 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–Y and 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}).

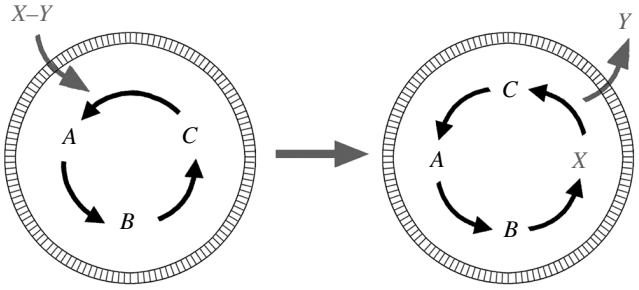

Figure 3. The foreign substance X–Y enters into the autopoietic structure and is capable of being incorporated into the existing metabolism, then modifying it.

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

[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.

in figure 1, there is no cognition in the terms explicated here, 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.

(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:

(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 self-maintenance (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.

REFERENCES

Bachman, P., Luisi, P. L. & Lang, J. 1992 Autocatalytic self-replicating 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://philsci-archive.pitt.edu/archive/00000888/index.html.

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 Schrödinger 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 de 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/rpl-l2h.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*. New York: North-Holland/Elsevier.
Varela, F. 1989*a* Reflections on the circulation of concepts between a biology of cognition and systemic family therapy. *Family Process* **28**, 15–24.
Varela, F. 1989*b* *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 self-reproduction 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.