

# **Enactive Cognitive Science and Biology of Cognition: A Response to Humberto Maturana**

*Tom Froese<sup>1, 2, 3</sup> and John Stewart<sup>4</sup>*

## **Introduction**

We very much appreciate that Maturana (2011) responded to our article, where we had made an attempt to excavate some of the hidden conceptual context in which the idea of autopoiesis had originally been formulated (Froese & Stewart, 2010). Our investigation was motivated by the growing interest in autopoiesis and related ideas among a new generations of researchers in cognitive science, driven by the increasing popularity of the enactive approach to cognitive science (Stewart, Gapenne, & Di Paolo, 2010). This enactive paradigm has been developed as an alternative to the traditional cognitivist-computationalist paradigm, and it is remarkable for its serious consideration of first-person experience and biological autonomy, two important domains of human existence that have so far been neglected in cognitive science.

Some proponents of the enactive paradigm have proposed that the concept of autopoiesis can serve as a potential explanatory link between the domains of phenomenological embodiment and biological embodiment. As a part of this project, we have recognized a need to go beyond considering living merely in abstract terms, such as an Ashbyan cybernetics of stability, and recognize that being alive essentially entails a precarious existence, or, as Maturana puts it, recognize “living beings as beings that die” (Maturana, 2011, p. 145). Given that only autopoietic systems are systems that can die, there is a possibility to ground the phenomenon of subjective existential concern in the objective process of autopoiesis (Weber & Varela, 2002). This is because only mortal beings can be concerned about their existence and therefore value its continuation and realization. Non-autopoietic systems persist, but they do not exist. In this way autopoiesis has provided the enactive paradigm with a useful conceptual locus in order to overcome Cartesian mind-matter dualism in a scientifically respectable manner (Thompson, 2007).

However, it has slowly become evident that the way in which autopoiesis had been described, following the primary literature of Maturana and Varela (1980), was inadequate for its new role in enactive cognitive science. For example, because

- 
1. Departamento de Ciencias de la Computación, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México (UNAM), México. Email: t.froese@gmail.com
  2. Centro de Ciencias de la Complejidad (C3), Universidad Nacional Autónoma de México (UNAM), México
  3. Ikegami Laboratory, Department of General Systems Studies, University of Tokyo, Japan
  4. Cognitive Research and Enaction Design (CRED), Université de Technologie de Compiègne, France

autopoiesis is usually taken to be an all-or-nothing class category, it left no conceptual room for explaining the possibility of gradations of adaptedness and varieties of concern (Di Paolo, 2005). For instance, sometimes we do well at managing our livelihoods, sometimes we fail our goals; sometimes people are healthy, sometimes they fall ill; sometimes people are brimming with life, sometimes they are at the brink of death. We experience these different kinds of existential situations faced with relevant concerns that make specific demands on our ability to interact with the world, but in all these cases we are equally autopoietic. Thus, there must be other processes going on inside the organism that can help us to explain these phenomena. As Maturana acknowledges, “The form of the relation of adaptation changes continuously in the realization of the living of the living being while its relation of adaptation is conserved” (Maturana, 2011, p. 148, footnote 4). But what kind of internal process regulates the continuous changes in the relation of adaptation such that the relation is conserved? How is their success evaluated such that we feel accordingly? Di Paolo (2005) proposed the concept of adaptivity to partially account for this process; living is no longer only autopoiesis, but living is autopoiesis and adaptivity.

Another related worry about the adequacy of the traditional description of autopoiesis for the enactive paradigm relates to the notion of *precariousness*. This notion, which derives from the existential biology of Jonas (2001), is in agreement with Maturana’s observation that living beings are beings that can die. In biological terms, dying is the process whereby an organism no longer adequately regulates its material and energetic exchanges with its immediate environment, such that its identity eventually disintegrates into the thermodynamic equilibrium. Conversely, we can say that living is a process of recursively maintaining one’s integrity as a dissipative structure in far-from-equilibrium conditions (Bickhard, 2009). Note that the autopoiesis of the living, which Maturana describes as “their continuous self production as discrete self producing dynamically closed molecular entities” (Maturana, 2011, p. 148), follows directly from this thermodynamic account of living and dying. On this view, autopoiesis is necessarily a physical process, and this process cannot be divorced from the adaptive regulation of material and thermodynamic interactions with its environment (Barandiaran & Moreno, 2008).

In this way the enactive paradigm tries to clarify the concrete phenomenological implications and concrete biological realizations of autopoiesis. Varela (1995) has referred to this goal of returning to the phenomena themselves as the re-enchantment of the concrete. Given this project, the important point we tried to make in our article was the following. Once we reject the traditional cybernetic conception of autopoiesis, which explicitly abstracts autopoiesis from its material and energetic realization (what we called the Ashbyan interpretation), and recognize that the concrete thermodynamic embodiment of the living is an essential aspect of autopoiesis, the concept’s current shortcomings for grounding our concrete phenomenological embodiment can be resolved. There is no longer any need to supplement autopoiesis with a shopping list of other conditions, such as adaptivity or precariousness. Similarly, there is no longer

any need for Maturana (2011, p. 144) to specify autopoiesis as “molecular autopoiesis,” as if to contrast it with non-molecular autopoiesis, because on the view we are defending there is no other kind of autopoiesis than a physically embodied one. Of course, on this view the concept of autopoiesis remains an observer’s abstraction of the specific way in which living beings realize their living, but it is an abstraction that takes more aspects of the biochemical context into account as constitutive of the phenomenon of the living.

Having briefly clarified the proposal of our previous article, we welcome the opportunity to continue this dialogue with Maturana by offering some further comments on various aspects of his response. Our aim is to highlight the possibility for developing a mutually informative and supportive exchange between the biological tradition of Maturana and his colleagues with the growing community of researchers interested in the paradigm of enaction. In particular, we try to show how Maturana’s approach in many ways complements the enactive critique of the cognitivist-computationalist paradigm of cognitive science, but we also point to some apparent differences. It remains to be seen whether these differences are superficial, perhaps terminological issues, or whether they are symptoms of more fundamental disagreements.

## **The Relevance of Biology of Cognition for Enactive Cognitive Science**

### *The Rejection of the Transcendental Subject*

Maturana makes the important point that we should not confuse our abstract theories with the concrete phenomena they are meant to denote. We agree that we need to be careful so that we do not unwittingly project some of our theoretical elements into the experiential domain in which we happen to distinguish the phenomena. As Maturana says, the phenomenon of living as such does not need any theory to ensure its existence.

We human beings can make theories about the nature of life when we think that life is some property of living beings, but life is not a property of living beings, the word life only evokes or names an invented abstract entity that we claim that must be there to sustain the living of a concrete singular living being. Living does not need any theory to occur; it is the occurring of a molecular autopoietic system. (Maturana, 2011, p. 146)

This is of course a general claim that is not specific only to the process of living. To give an example from physics, planetary movement does not need Newton’s equations of gravity to occur, it is the occurring of a gravitational system. More specifically, although it is common practice to loosely talk about gravity constituting planetary motion, it would strictly speaking be more accurate to say that the concrete phenomenon, which we conventionally happen to denote and describe by the abstract concept of gravity, is what constitutes planetary motion. Similarly, we can say that the phenomenon of living, which can be described as a molecular autopoietic system, does not need autopoietic theory to occur. For most practical purposes it may well be

too cumbersome to spell out scientific linguistic references in this more precise manner. However, if we are not careful about maintaining the essential distinction between a concrete phenomenon and its abstract description, then it may eventually happen that the description unwittingly takes the place of the phenomenon. In the worst-case scenario we can get caught up in a fruitless debate about something, which is in fact nothing but a figment of our imagination (more on this later).

The reverse can also happen, whereby figments of our imagination without any empirical foundation are confused as scientific explanations of concrete phenomena. Maturana may have unfortunately misunderstood our appeal to what we called a Kantian interpretation of autopoiesis in this manner. However, this interpretation is not a reference to Kant's concept of a transcendental subject, which is conceived as an independent a priori organizing principle, and we therefore agree with Maturana's claim that "when one speaks of the autonomy of living systems one is saying that their operation as discrete living entities follows regularities determined by the manner they are made, and *not by any external organizing factor*" (Maturana, 2011, p. 143; emphasis added). Although we cannot go into more details here, our Kantian interpretation of autopoiesis was not intended as an attempt to turn autopoiesis into an external transcendental principle. On the contrary, it was based on Kant's observation that all living beings are self-organizing (he introduced this word) and produce their own components (Weber & Varela, 2002). Kant argued that it is this unique being of the living, which may help us to explain why organisms appear to us as purposeful beings in the relational domain in which we encounter them. Of course, artifacts have a purpose, too, but even in the case of robots the purpose is only derivative of the goals of their maker and designer (Froese & Ziemke, 2009). Organisms also have a purpose, but in that case the purpose is intrinsic to their existence as living beings, because as autopoietic systems they are their own maker (self-producing) and designer (self-organization).

In this way the enactive account of the phenomenon of living tries to chart a course between the idea of a transcendental subject on the one hand (vitalism) and the cybernetic rejection of all intrinsic purpose on the other (nihilism). Autopoiesis is crucial to this approach because it describes how a process can recursively specify its own identity and conditions of operation.

### *The Rejection of the Representational Theory of Mind*

Mainstream cognitive science has also tried to navigate a path between vitalism and nihilism, but they have done so by attributing their abstractions from personal-level psychology to the states of the component-level mechanism. According to this view, mental states are realized by inner mental representations. The enactive approach and Maturana's biology of cognition are united in rejecting this representational theory of mind. Given the ubiquity of this kind of theory in cognitive science and biology, we appreciate Maturana's insistence on the principle of non-intersecting domains. Maturana describes this principle in the following manner:

All systems are composite entities that exist in two not intersecting operational-relational domains, the domain of the operation of their components, and the domain of their operation as totalities. Due to this the totality does not operate as an argument in what happens with its components, and the components do not operate as arguments in what happens with the totality. (Maturana, 2011, p. 147)

The basic idea is that we always need to be aware of the level of description at which we are working, and not to mix elements across different levels. For instance, there is a fundamental difference between interacting with a computer as a composite system of different circuit components, and interacting with a computer as a whole system. A property that pertains to the computer as a whole, for instance that it is running a word processing program, does not entail that an internal component has that same property. Similarly, a property that pertains to a single component of the computer, for example that a bit switch is limited to be in one of two states, does not entail that the computer as a whole has that same limiting property.

Maturana's principle of non-intersecting domains may seem trivially obvious, but in fact most of cognitive science has been built by systematically violating it (Froese, 2010). The very idea of the cognitive unconscious is based on the two-step approach of (1) distinguishing some properties and abilities that pertain to a situated person as a whole, such as mental states and thinking, and (2) re-assigning those properties and abilities to states and modules inside that person's brain. This fallacy is surprisingly difficult to eradicate (Searle, 1990). And the same tragedy has happened in modern biology, where traits of the organism-niche as a whole have been re-assigned to the internal genetic system, an approach that continues to be popularized as the selfish gene. Again, this fallacy is proving to be surprisingly resilient (Rose, 2005).

Maturana is therefore correct in highlighting the need to distinguish between different levels of observation and explanation, especially regarding the crucial difference between a system as a unity and as a system of constituting components. And in this context he is also right that we need to be careful when we say that living beings are purposeful beings. We clearly do not want to perpetuate the mistakes of researchers who, in their attempt to explain what they experience in their lives, are "resorting to some ordering principle that acts in a manner comparable to how they act with purposeful intent in ordering their daily affairs" (Maturana, 2011, p. 148). There is no cognitive homunculus inside the brain. As Ashby (1960) demonstrated with his concept of ultrastability, just because a system as a whole behaves intelligently for an observer, this does not entail that there is an intelligent component on its inside. Similarly, let us be clear that when we talk about purpose, meaning, value, and so forth we are not trying to add reified versions of these concepts into an agent's level of components. According to the paradigm of enaction these concepts refer to qualities that pertain to the relational domain of agent-environment interaction (Di Paolo, Rohde, & De Jaegher, 2010).

### *The Rejection of Internalist Theories of Mind*

Following on from this relational account of intentionality, here again we side with Maturana against the prevailing internalist doctrine of cognitive science and biology.

Although some interpretations of the literature on autopoiesis may give the impression that the rejection of the representational theory of mind is supposed to make space for an alternative kind of internalist theory of mind (Wheeler, 2010), this is not the way in which the enactive approach is currently developing. The phenomena that we usually denote as life, mind, or cognition are relational phenomena that pertain to an embodied agent as a whole in the context of its niche (Froese & Di Paolo, 2011). As human beings we are not trapped inside in our heads or bodies, as the internalist approach assumes. And neither can we somehow step outside of the relational domain that constitutes our living existence, as the transcendental approach would require us to do. The relational domain is the domain in which all of our activity takes place.

All that we can say or distinguish in us in our operation as observing, reflecting, languaging, self-conscious human beings, as we operate as reflective persons, occurs in our operation as persons in the relational space that arises in our operation as totalities as molecular autopoietic systems, and not in the molecular dynamics of the realization of our autopoiesis. (Maturana, 2011, p. 144)

This kind of relational description of the mind offers a valuable alternative to the traditional internalism of cognitive science, which remains trapped in the head even in its attempt to argue for extended cognition (Di Paolo, 2009). However, at the same time in the context of this general agreement on a relational approach, there is scope for disagreement about its precise details. For example, we are not convinced by Maturana's restriction of significance to the worlds of human beings with language.

Purpose, meaning, value, utility, advantage, importance, regulation, cause, control ... are reflexive commentaries that an observer makes about what happens in the human sensory-operational-relational space, and are not features or characteristics of any process. (Maturana, 2011, p. 147)

Why should other animals not also live in worlds that are meaningful to them? Of course, if when Maturana claims that "sense and meaning belong to the domain of the commentaries and reflections that an observer makes about what he or she sees" (Maturana, 2011, p. 147), he is merely claiming that the concepts of sense and meaning are part of our linguistic domain and not part of the relational domain of other non-linguistic animals, this is fine. Similarly, the concept of gravity *as a concept* is not making the moon move around the earth. But the important point is that these concepts are intended to connote phenomena that can be distinguished in our lived experience, and it is these phenomena that are ultimately of interest. For example, when we are engaging with non-linguistic animals we perceive their own relational domains as being meaningful for them, and this phenomenon deserves an explanation that takes the lived perspective of the animal into account. As the biologist von Uexküll (1957) has demonstrated with many examples, it is possible to devise richly detailed descriptions of the meaningful relational worlds of other animals.

We are not sure if Maturana would deny that animals live in meaningful worlds, especially as his notion of languaging is a relatively broad. What is certain is that the paradigm of enaction is making efforts to overcome the old Cartesian metaphysical

dualism between human beings as spirit-like observers and other living beings as mere clockwork mechanisms. This human-centered dualism ignores the evidence of our own lived experience, and the biological and historical evidence of continuity among all living beings, which share common principles and arose from a common ancestor (Jonas, 2001).

Part of the problem here may be Maturana's overgeneralization of the abstract stance of the biochemist. For instance, in response to his question regarding the origins of living beings, he claims that we must look at living systems now and goes on to say:

And the answer is pointing to *what we all can see*, namely, a living system is a discrete self-contained molecular dynamic system that produce itself as a closed network of productions of molecules that in their interactions produce the same network of molecular productions that produced them as a stationary dynamics sustained in a continuous flow of matter and energy through it. (Maturana, 2011, p. 145; emphasis added)

When we look at a living being, be it another fellow human being or an animal, like a pet, is this really what we all can perceive? Is this what we experience when we look at ourselves, for example our face in the mirror or the flesh of our arms? Even when we happen to see the components inside a living being, perhaps because it is being dissected, it is still unlikely that we *see in* its fleshy mess such abstract things as molecules, stationary dynamics, or a flow of energy. Unless Maturana is actually living an experience that we, and presumably most other people, do not share, this seems to be a confusion of experiential and explanatory domains. Perhaps Maturana is speaking of looking and seeing in a metaphorical way only, and he does not intend to comment on his experience. But if Maturana does not mean to say that this is how he perceives living beings, then how do they appear in his experience instead? What marks them as living in his perspective? Do they appear as a locus of agency or as the same kind of material being as, for example, a rock? Do they appear as static or as always in flux and as becoming different (e.g., growing, developing, dying, etc.)?

We simply insist on Maturana's point that we need to be very careful to distinguish between how living beings normally appear to us in our experience, and how we describe them and explain them. We want to ensure that we are trying to describe and account for the same kind of phenomenon. For us, at least, other animals appear to us as being situated in meaningful worlds of their own, as long as they are perceived as a whole. Organisms are clearly non-stationary phenomena that exhibit goal-directed behavior, growth, development, and ageing. When dissecting organisms and studying the organization of their components at the molecular level, we can abstract the notion of an autopoietic system by adopting an appropriate theoretical stance. But this stance requires that we have already identified a whole organism to begin with; it cannot be the starting point. And there are lots of different levels of distinctions that we can make (e.g., molecular, chemical, cellular, physiological). These levels of description should not be conflated with each other, and we should try

to resist moving too quickly from the phenomenon of living to the idea of a living system at the molecular level.

*The Appeal to Conceptual and Methodological Clarity*

Maturana rightly insists that we must always remain vigilant of not falling into linguistic traps when practicing the scientific method. He repeatedly warns us to avoid the “confusion of experiential and explanatory domains” (Maturana, 2011, p. 151), and he reminds us that “one formalizes what one thinks about the processes that one wants to formalize and not the process one believes to be formalizing” (Maturana, p. 143). We agree with the importance of these distinctions. In the philosophy of science they can be traced back to the work of Hempel and Oppenheim:

We divide an explanation into two major constituents, the explanandum and the explanans. By the explanandum, we understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the explanans, the class of those sentences which are adduced to account for the phenomenon. (Hempel & Oppenheim, 1948, pp. 136-137)

In other words, the practice of generating a scientific explanation involves three distinct elements, namely a concrete *phenomenon* that we have distinguished in our experience, an abstract description of that phenomenon (the *explanandum*), and an explanatory account of that explanandum (the *explanans*). Ideally, scientists should proceed with a clear awareness of the essential differences between these three elements. However, in practice the distinctions are all too often forgotten, which can have disastrous consequences.

For instance, for several decades there has been an ongoing debate in cognitive science about the so-called theory of mind. The idea of a theory of mind first began as a philosophically motivated account of social understanding, that is, of how people manage to make sense of each other (explanans). But at the same time the very phenomenon of how people manage to understand each other was described in terms of a theory of mind (explanandum), and it was simply taken for granted that this explanandum indeed matched the phenomenon of how people manage to understand each other (Ratcliffe, 2007). Consequently, no clear distinction has been made between the phenomenon, its description and its account. What will happen if the enactive approach is correct in insisting that the description does not match the target phenomenon after all (Froese & Fuchs, 2012)? What if, for instance, social understanding is actually more accurately described as a matter of mutual interaction and languaging? In that case most of the debate about so-called social cognition that has been going on for the last 50 years will turn out to be mostly useless, no matter how refined the accounts (explanans) of the description (explanandum) have become over the years.

What this example shows is that the practice of conceptual and methodological clarity must be supplemented by phenomenological clarity, and this insight was one major motivation for Varela’s enactive approach to cognitive science (Varela, Thompson, & Rosch, 1991). To be fair, Maturana has also highlighted the pre-

theoretical praxis in which we always already find ourselves situated, but he has not undertaken any systematic phenomenological exploration of this domain or grappled with the methodological problems of how to do so (Froese, 2011b). In contrast, Varela spent much of his later career formulating pragmatic methods that help us to better become aware of our experiencing (Depraz, Varela, & Vermersch, 2003). These kinds of first-person methods are slowly finding their way into cognitive science (Froese, Gould, & Barrett, 2011). Phenomenology is one of the areas where the biology of cognition could learn from the enactive approach.

### **The Relevance of Enactive Cognitive Science for Biology of Cognition**

One motivation of our original article was the worry that the debate about autopoiesis had run into trouble. As we already summarized in the introduction, there were incompatibilities when trying to integrate autopoiesis into an understanding of life and mind that paid closer attention to the phenomenology. One aim of our article was therefore to evaluate whether the abstract description of living beings on the level of components (autopoiesis) was adequate to the concrete phenomenon, that is, if autopoiesis was in fact a sufficiently accurate description of the material realization of a living being (explanandum).

However, Maturana's response to our article has made us unsure if we understood correctly what he means by the term *autopoiesis*. In this final section we therefore highlight why we are uncertain about our scholarly interpretation, and we also outline our suggestion of how to integrate the notion of autopoiesis into a more progressive research program. As we clarified above, the enactive approach is sympathetic to many of Maturana's central insights. We wish to conclude this commentary by indicating how those insights can be developed further.

#### *Autopoiesis: Phenomenon, Explanandum, or Explanans?*

We are surprised about Maturana's remark that we "speak as if [we] thought that the notion of autopoiesis were something that existed out there in the world and about which one can have different opinions" (Maturana, 2011, p. 143). In what sense does he disagree with this way of speaking? Of course, we do not claim that the concept of autopoiesis *as a concept* exists out there in the world, other than as a part of our practice of languaging, but even then we can definitely have different opinions about it. Surely Maturana does not believe that autopoiesis is some kind of transcendental principle that stands outside the world and is therefore beyond debate? We certainly do claim, and we hope Maturana concurs, that we can have different opinions about anything of relevance; to claim otherwise is to reject the possibility of dialogue and therefore of science (not to say anything of more sinister possibilities). In the context of biology this means that the notion of autopoiesis is involved in an open debate about what phenomenon to distinguish as the living, how to describe it, and how to explain it.

Accordingly, the first fundamental issue that must be clarified is whether the concept of autopoiesis is typically used to refer to a concrete phenomenon, an abstract description of that phenomenon (explanandum), or an explanatory account of that description (explanans). Given all that has been said above about the need for conceptual and methodological clarity, this may seem like an easy task. But the problem is that Maturana appears to employ the term *autopoiesis* for all three of these roles. At some points it sounds like he uses autopoiesis as a way of explaining why there are qualitative differences between living systems and other kinds of dissipative systems, that is, that autopoiesis is part of an explanans.

living systems as dissipating systems are different from candles and tornados, and the difference is that they are autopoietic systems, candles are candles and tornados are tornados. Yes, what makes any kind of dissipating system different from another kind of dissipating system is the organization that *defines* their different class identities as dissipating systems of a different kind. (Maturana, 2011, p. 146; emphasis added)

As we have discussed, the enactive approach is also interested in using autopoiesis as part of an explanatory account of why relational domains of living beings are qualitatively different from the relational domains of other beings. But given that Maturana also makes it clear that the organization of a system *defines* its class identity, it may be more appropriate to interpret him as claiming that the autopoietic organization is an *a priori* logical category that includes all living systems, no matter what research will uncover about them. This is indeed a common interpretation of autopoiesis in the secondary literature, and it would entail that autopoiesis is dependent on how we desire to carve up the space of systems, rather than an explanatory account or a description of an empirical phenomenon.

But again Maturana may disagree with this interpretation, as he also says that “autopoiesis is *not* a definition, not a postulate or an *a priori* proposition, it is an *abstraction* of what happens in the actual realization of the living of a living system” (Maturana, 2011, p. 146; emphasis changed). According to this statement, it seems that the term *autopoiesis* is actually part of an abstract description of the phenomenon of the living, that is, that autopoiesis is part of an explanandum. On our view, this is the most fitting role for autopoiesis, namely that it is a label for an abstract description of the kinds of circular processes of self-production that characterize living beings when they are studied as systems of components. It is in this spirit that we proposed to modify the way autopoiesis has been traditionally defined so as to make it less abstract and more specific to the concrete phenomenon by including thermodynamic conditions in its description (Froese & Stewart, 2010).

However, there is still a third possibility for the role of autopoiesis. As we indicated in our discussion above, Maturana’s remark that when we “look” at a living being “we all can see” that “a living system is a discrete self-contained molecular dynamic system” (Maturana, 2011, p. 145) seems to imply that autopoiesis is actually something that we can distinguish in our experience, that is, that autopoiesis is part of, or perhaps even identical to, the concrete phenomenon of living. As Maturana himself

puts it at one point, “the molecular autopoiesis of a cell *is* its living” (Maturana, p. 144; emphasis added). However, from an enactive point of view, this ontological interpretation of autopoiesis is not an attractive option. As we observed above, there is in fact much more to the experienced phenomenon of living than what the term autopoiesis is meant to pick out (Jonas, 2001). Then there is the worry that molecules are not part of the phenomenon at all, but rather explanations that are proposed by an objectivist theoretical stance (Stewart, 2001).

Given these confusing and conflicting uses of the term *autopoiesis*, we can better understand why there has been some reluctance in mainstream cognitive science and biology to engage with this term. As we remarked above, on our view the most reasonable interpretation is that autopoiesis is a shorthand label for a description of the circular processes of self-production that are essential to all living beings in as much as they depend on these processes to exist. To put it crudely, autopoiesis is another word for what biologists describe as metabolism. But if this is the case, then it is also easier to see why it is often difficult to understand what all the fuss about autopoiesis is about (e.g., Dennett, 2011). What do we gain by saying autopoiesis rather than metabolism? Perhaps this change in terminology enables us to better highlight that circular processes of self-production provide an organism with an intrinsic autonomy, and this autonomy plays an important explanatory role in accounting for the relational phenomenon of meaningful and intentional agency (Barandiaran, Di Paolo, & Rohde, 2009).<sup>5</sup>

#### *From Non-intersecting Domains to Co-dependent Realization*

Notice, however, that an explanation of the phenomenal quality of an organism’s relational domain in terms of the internal organization of the components of that organism requires a loosening of Maturana’s principle of non-intersecting domains. For if (1) there is absolutely no intrinsic relationship between a unity and its components, and (2) autopoiesis pertains to the organization of the components only, and (3) behavior, cognition, and lived experience pertain to the relational domain of the unity only, then (4) it logically follows that the concept of autopoiesis is useless for the science of mind. On this view, the biology of cognition would be forced to accept an alternative kind of mind-body dualism (Di Paolo, 2009).

One exciting aspect of the paradigm of enaction is that it rejects the conclusion (4) by arguing that (1) must be relativized somehow, while continuing to insist on the validity of (2) and (3). This approach certainly does not necessitate a return to some kind of representationalist or internalist theory of mind, and it also means that we have to remain aware of the distinctions that we make as observers. In fact, precisely because it is we as observers who are making the explicit distinction between the

---

5. With regard to the constitution of meaning by the organism, the biological foundations of the enactive approach are at odds with the claim, typical of the traditional biology of cognition, that living systems are purposeless systems, although there may be a resolution to this difference (Thompson, 2004). The enactive approach seems more related to biosemiotics (e.g. Emmeche, 1998; Hoffmeyer, 2000), but a scholarly comparison between these two traditions is still missing.

phenomenon as a unity and the phenomenon as a system of components, we know that this is not a distinction that pertains to that phenomenon as such. Therefore, neither is the principle of non-intersecting domains an intrinsic aspect of a living being. On the contrary, a living being is one existence as a whole, with which we can interact in different ways, including by directly interacting with its insides (e.g., if we are a doctor or veterinarian). Similarly, we do not experience ourselves to be absolutely independent from our living body, rather we live our body, and it is this concrete living and lived embodiment which unifies our constitutive and relational domains of existence (Froese, 2011a). This is the enactive notion of the embodied mind (Thompson, 2007). In other words, while we agree that we need to remain aware of our role as observers, we argue that the different stances that we can adopt to study the phenomenon of living (e.g., by treating a living being as a system of molecules, or chemicals, or cells, or organs, etc.) pick out different co-dependent aspects of one coherently integrated living being (Rose, 2005).

We can start to establish links between the different descriptive levels when we consider more concretely their co-dependent realization. For instance, the relational domain of a dissipative structure such as a candle or a tornado is characterized by its self-maintenance in far-from-equilibrium conditions. And it is this far-from-equilibrium condition, which drives the processes that constitute the dissipative structure such that it can maintain itself. In reaction-diffusion systems we can find individuated dissipative structures, which can be considered examples of molecular autopoietic systems, that exhibit self-motility due to different rates of growth and decay (Froese, Virgo, & Ikegami, 2011). The constitutive and the relational domain can be distinguished, but they cannot be absolutely dissociated.

Living systems take this principle of co-dependent realization further by adaptively regulating their internal processes and relational interactions (Barandiaran & Moreno, 2008). This recursive self-maintenance clearly distinguishes them from other kinds of dissipative systems, perhaps even more so than their autopoiesis (Bickhard, 2009). For example, all organisms will normally regulate their interactions in search of nutrients and move away from danger, but we do not expect a candle to go looking for more wax or to seek shelter from the wind. Cognition is a form of such regulation (Froese & Di Paolo, 2011). Moreover, many regular behaviors are integrated within the existence of living beings to such an extent that they are a necessary aspect of the self-constitution of the organism. This is certainly the case for basic behaviors (e.g., breathing, eating, drinking, excretion), but it can also apply to acquired behaviors (e.g., habits, addictions) and social interaction (e.g., babies require contact to survive and to develop). The being and doing of living are intertwined. Thus, rather than following Maturana in identifying the process of living with the autopoiesis of internal components, we can begin to extend autopoiesis beyond the cell membrane to incorporate processes in the ecological context (Virgo, Egbert, & Froese, 2011) and speak of a relational or extended life (Di Paolo, 2009). It is no longer straightforward to distinguish self from other or insides and outsides

(Hoffmeyer, 1998). On this view, we are always already entangled with the world and participate in each other's lives (Froese & Fuchs, 2012).

## An Opening for Further Discussion

We are convinced that the concept of autopoiesis has a valuable role to play in the enactive approach to cognitive science. Although we argued that the concept must be further clarified, we hope to have shown that the enactive approach is generally sympathetic to many central insights of Maturana's biology of cognition. In particular, it agrees with the criticisms against the mainstream internalist-representationalist theories of life and mind, and it therefore strives to develop an alternative theory that better takes our biological existence into account. At the same time it recognizes our pre-reflective embodiment in what Maturana calls our praxis of living, and it aims to integrate phenomenology and science in order to ensure that we always remain aware of the distinction between phenomenon and theory. Much more can be said about all of the issues we discussed in this paper, and it will be interesting to determine what, if any, are the fundamental points of disagreement. We highlighted some important differences as an opening for further mutually informing debate.

## References

- Ashby, W. R. (1960). *Design for a brain: The origin of adaptive behaviour* (rev. ed.). London: Chapman & Hall.
- Barandiaran, X., Di Paolo, E. A., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior*, 17 (5), 367-386.
- Barandiaran, X., & Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive Behavior*, 16 (5), 325-344.
- Bickhard, M. H. (2009). The biological foundations of cognitive science. *New Ideas in Psychology*, 27, 75-84.
- Dennett, D. C. (2011). Shall we tango? No, but thanks for asking. *Journal of Consciousness Studies*, 18 (5-6), 23-34.
- Depraz, N., Varela, F. J., & Vermersch, P. (2003). *On becoming aware: A pragmatics of experiencing*. Amsterdam: John Benjamins Publishing.
- Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4 (4), 429-452.
- Di Paolo, E. A. (2009). Extended life. *Topoi*, 28 (1), 9-21.
- Di Paolo, E. A., Rohde, M., & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne & E. A. Di Paolo (Eds.), *Enaction: Toward a New Paradigm for Cognitive Science* (pp. 33-87). Cambridge, MA: The MIT Press.
- Emmeche, C. (1998). Defining life as a semiotic phenomenon. *Cybernetics & Human Knowing*, 5 (1), 3-17.
- Froese, T. (2010). From cybernetics to second-order cybernetics: A comparative analysis of their central ideas. *Constructivist Foundations*, 5 (2), 75-85.
- Froese, T. (2011a). Breathing new life into cognitive science. *Avant. The Journal of the Philosophical-Interdisciplinary Vanguard*, 2 (1), 95-111.
- Froese, T. (2011b). From second-order cybernetics to enactive cognitive science: Varela's turn from epistemology to phenomenology. *Systems Research and Behavioral Science*, 28, 631-645.
- Froese, T., & Di Paolo, E. A. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition*, 19 (1), 1-36.
- Froese, T., & Fuchs, T. (2012). The extended body: A case study in the neurophenomenology of social interaction. *Phenomenology and the Cognitive Sciences*, 11 (2), 205-235.
- Froese, T., Gould, C., & Barrett, A. (2011). Re-viewing from within: A commentary on first- and second-person methods in the science of consciousness. *Constructivist Foundations*, 6 (2), 254-269.
- Froese, T., & Stewart, J. (2010). Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. *Cybernetics & Human Knowing*, 17 (4), 83-106.
- Froese, T., Virgo, N., & Ikegami, T. (2011). Life as a process of open-ended becoming: Analysis of a minimal model. In T. Lenaerts, M. Giacobini, H. Bersini, P. Bourguin, M. Dorigo & R. Doursat (Eds.), *Advances in Artificial Life, ECAL 2011: Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems* (pp. 250-257). Cambridge, MA: The MIT Press.

- Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, 173 (3-4), 366-500.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15 (2), 135-175.
- Hoffmeyer, J. (1998). Surfaces inside surfaces. On the origin of agency and life. *Cybernetics & Human Knowing*, 5 (1), 33-42.
- Hoffmeyer, J. (2000). The biology of signification. *Perspectives in Biology and Medicine*, 43 (2), 252-268.
- Jonas, H. (2001). *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston, IL: Northwestern University Press.
- Maturana, H. R. (2011). Ultrastability ... autopoiesis? Reflective response to Tom Froese and John Stewart. *Cybernetics & Human Knowing*, 18 (1-2), 143-152.
- Maturana, H. R., & Varela, F. J. (1980). Autopoiesis: The Organization of the Living *Autopoiesis and Cognition: The Realization of the Living* (pp. 59-140). Dordrecht, Holland: Kluwer Academic.
- Ratcliffe, M. (2007). *Rethinking commonsense psychology: A critique of folk psychology, theory of mind and simulation*. New York: Palgrave Macmillan.
- Rose, S. (2005). *Lifelines: Life beyond the gene*. London: Vintage Books.
- Searle, J. R. (1990). Consciousness, explanatory inversion, and cognitive science. *Behavioral and Brain Sciences*, 13, 585-642.
- Stewart, J. (2001). Radical constructivism in biology and cognitive science. *Foundations of Science*, 6 (1-3), 99-124.
- Stewart, J., Gapenne, O., & Di Paolo, E. A. (Eds.). (2010). *Enaction: Toward a new paradigm for cognitive science*. Cambridge, MA: The MIT Press.
- Thompson, E. (2004). Life and mind: From autopoiesis to neurophenomenology. A tribute to Francisco Varela. *Phenomenology and the Cognitive Sciences*, 3 (4), 381-398.
- Thompson, E. (2007). *Mind in Life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: The Belknap Press.
- Varela, F. J. (1995). The Re-enchantment of the concrete: Some biological ingredients for a nouvelle cognitive science. In L. Steels & R. Brooks (Eds.), *The artificial life route to artificial intelligence* (pp. 11-22). Hove, UK: Lawrence Erlbaum Associates.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: The MIT Press.
- Virgo, N., Egbert, M. D., & Froese, T. (2011). The role of the spatial boundary in autopoiesis. In G. Kampis, I. Karsai & E. Szathmáry (Eds.), *Advances in artificial life: Darwin meets von Neumann. 10th European Conference, ECAL 2009* (pp. 234-241). Berlin, Germany: Springer-Verlag.
- von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed.), *Instinctive behavior: The development of a modern concept* (pp. 5-80). New York: International Universities Press.
- Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1, 97-125.
- Wheeler, M. (2010). Minds, things, and materiality. In L. Malafouris & C. Renfrew (Eds.), *The cognitive life of things: Recasting boundaries of the mind*. Cambridge: McDonald Institute for Archeological Research.



Young, I. (2012). *Fairing the Hull*. Unretouched photograph.