# Extended life

Ezequiel Di Paolo

Centre for Computational Neuroscience and Robotics
Centre for Research in Cognitive Science
University of Sussex
Brighton, BN1 9QH, UK
Tel: (+44) 1273 877763
Fax: (+44) 1273 877873
E-mail: ezequiel@sussex.ac.uk

**Abstract.**

This paper reformulates some of the questions raised by extended mind theorists from an enactive, life/mind continuity perspective. Because of its reliance on concepts such as autopoiesis, the enactive approach has been deemed internalist and thus incompatible with the extended mind hypothesis. This paper answers this criticism by showing 1) that the relation between organism and cogniser is not one of co-extension, 2) that cognition is a relational phenomenon and thereby has no location, and 3) that the individuality of a cogniser is inevitably linked with the question of its autonomy, a question ignored by the extended mind hypothesis but for which the enactive approach proposes a precise, operational, albeit non-functionalist answer. The paper raises a pespective of embedded and intersecting forms of autonomous identity generation, some of which correspond to the canonical cases discussed in the extended mind literature, but on the whole of wider generality. In addressing these issues, this paper proposes unbiased, non-species specific definitions of cognition, agency and mediation, thus filling in gaps in the extended mind debates that have led to paradoxical situations and a problematic over-reliance on intutions about what counts as cognitive.

**Keywords:** extended mind, enactive approach, autonomy, precariousness, agency, mediation, individuation.

## 1. Introduction

In a recent article, Michael Wheeler (2008) has put a question mark on the relation between enactivism (in its life-mind continuity version, e.g., Varela, Thompson & Rosch, 1991, Thompson, 2007) and the extended mind (EM) hypothesis (Clark & Chalmers, 1998). His conclusion spells gloom for the prospects of a unified non-Cartesian cognitive science: enactivism and EM are demonstrably incompatible!

This conclusion seems at odds with the spontaneous understanding of enactivism as proposing a view of cognition as fundamentally embodied and situated and the (apparently!) parallel understanding of EM as signalling how much of our cognitive skills rely crucially on the availability of non-biological epistemic technologies. As someone interested in the enactive programme, my first reaction to this declaration of incompatibility has been to take Wheeler's argument as a much-needed call for clarification. Surely, something is wrong in how elements of the enactive approach have been expressed that triggers him to follow an internalist line of interpretation? Surely, it should be possible to clarify these issues and redraw the conceptual boundaries with a thicker pen in order to avoid such unintended readings? And surely the end of such an exercise would be a happy re-unification in the promised land of non-Cartesianism?

As it turns out, Wheeler's conclusion is right, but not for the reasons he puts forwards. And to spell out the reasons why I think he's right will reveal a radically different assessment of the situation. To be sure, it will be necessary to point to what I consider misleading interpretations of what the enactive approach stands for (sometimes the product of ambiguities in the main literature). But I think that a more urgent task will be to signal the degree to which the EM is a symptom of the blind spots of traditional cognitive science (understood as cognitivist accounts of the mind that are largely representational and functionalist). For doing this, it will be important to ask again some of the questions posed by EM theorists from an enactive perspective. In raising such questions, we shall make clear two important aspects of enactivism: the problems it is trying to solve as a research programme and how it attempts to address them. Already in the problems themselves we can start drawing some distinctions that separate the enactive programme from the EM as classically formulated, for these are precisely the questions that are *avoided* by the latter: questions about individuation, autonomy, agency, normativity, and the nature of cognition. In my view, this avoidance is pathological in the strict sense that it leads to ill-posed problems and creates a "moth-around-the-candle" effect. Thus, the EM hypothesis has the form of an unfinished *reductio ad absurdum*. Completing this *reductio* would mean to raise the questions that motivate the enactive perspective. Herein lies the incompatibility between the approaches. But as a positive opening among the clouds, the way that enactivism has begun to answer some of its questions provides novel tools and new conceptual distinctions. And these are, in my opinion, quite compatible with the spirit of EM, which I understand as non-bio-chauvinist and eager to unbind minds from skulls. For the enactivist, mind was never in-the-head in the first place.

## 2. Location, location, (and only) location

Before asking where it is we *must* first say what it is. This is the single major problem with the way EM theorists have approached the genuine question of whether extra-neural, extra-bodily material processes are a constitutive part of what we intuitively recognize as cognitive processes. Relying solely on those intuitions is the problem.

Consider the parity principle, which is a useful way to introduce what EM is all about:

> If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process (Clark and Chalmers, 1998 p. 8).

Already in this formulation we can see that there seems to be no question of providing a definition of what should count as a cognitive process in order then to pose the subsequent question of whether such processes are skull-bound, neuronal, bodily or extended. Clark even goes as far as to say that the parity principle "identifies a state or process as cognitive" (Clark forthcoming, manuscript 20). Only recently has the debate moved into considering this question that should have been its point of departure. The parity principle relies both on simple prejudices about inner and outer as well as on intuitions about cognition. EM wishes to question the first set of prejudices but, curiously, it remains silent about the second. The principle is formulated as the functional equivalence of a "worldly" process that "were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process". Instead of moving from the good case histories that lead to the parity principle into a questioning of our intuitions about boundaries *and* our intuitions about cognition, the EM proponents have strangely chosen to ignore the latter route and have only questioned the boundaries. But – and here we confront the pathology – the intuitions about cognition that EM relies on are inevitably tied to the boundaries between inner and outer that it wishes to undermine (witness the phrase "were to go on in the head" as the sole arbiter of whether our intuitions about cognition are correct).

The conflation of two different kinds of boundary is at play here. The boundary of the organism (or its brain) as the "intuitive" locus of what we refer to as a cognitive system, and the more abstract boundary between processes, engagements, relations, mechanisms and systems that deserve the name cognitive and those that do not. The second boundary is never explicitly thematized, but it is there.

> [The] organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in the system play an active causal role, and they jointly govern behavior in the same sort of way that cognition usually does. If we remove the external component the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head (Clark & Chalmers 1998, p 7).

This seems to suggest a practical and operational way out of the problem. Perform a causal analysis of the coupled system and work out what processes contribute to cognitive performance. But of course, without a measure of relevance, a causal analysis will inevitably invite an unbounded spread of causes (e.g., isn't oxygen *obviously* crucial for a human to solve math problems?). It is clear that what counts as cognitive (the second boundary) should be the measure that determines the *relevance* of the causal contribution of a given process. But this leads us again to the problem already stated. The only test of the cognitive offered by EM is whether we intuitively would call something cognitive were it to happen in the head. At the last, crucial stage, the conceptual boundary of the cognitive is regulated by the physical boundary of the skull in the very movement that questions the validity of this regulation.

Just imagine how an EM researcher would go about determining whether extra-bodily process X is constitutive of a cognitive process. He is at this point already convinced that the skull boundary no longer makes sense for determining whether something is cognitive. And yet, on consulting his EM handbook he finds this schizophrenic piece of advice: "Even though you know that you should not rely on the skull boundary in order to call process X cognitive, you must still use it to check whether it would have still made sense to those confused people who did not take EM seriously and consider how they would have judged process X were it to happen in the head". Isn't this the absurd equivalent of "Even though in the past several cases of mental illness were erroneously diagnosed as cases of demonic possession, in modern days, in order to determine whether you have a case of mental illness you must always consult an exorcist first"?

Notice that this is not the same line of argument against EM as that presented, for instance, by Rupert (2004). Accordingly, the benchmark of the cognitive would be given by the fine-grained specification of processes deemed functionally equivalent to external process X but which may ultimately reveal functional differences in detail. Wheeler is able to effectively deflect this line of argument (2008, forthcoming) by a laborious re-reading of the counterfactual element of the parity principle ("were it to happen in the head"). However, my critique is sitting at a different level, which is that of the very appeal to intuitions about cognition instead of a questioning of those intuitions. Clark states: "The parity principle was meant to engage our rough sense of what we might intuitively judge to belong to the domain of cognition—rather than, say, that of digestion—but to do so without the pervasive distractions of skin and skull" (Clark, 2007, p 166). It is a sort of "equal opportunities" policy. But what's determining the job to be done? A regulatory appeal to pre-theoretical intuitions about cognition. Even though for Clark it would seem that such intuitive notions of the cognitive are locationally neutral, this pre-theoretical understanding is biased toward the inner, as discussed in detail by Wheeler (forthcoming) and unavoidably suggested by the parity principle. What we lack, and enactivism is able to offer, is a theoretically loaded and locationally neutral account of cognition.

The *reductio* (the questioning of the assumptions that lead to the paradoxical situation of conceptually dissolving a boundary whilst conserving it for methodological purpuses) is never fully accomplished by the EM literature, though recently, following a similar path, Clark has raised the question of whether we should eliminate the concept of mind altogether (Clark, forthcoming). In doing so, he acknowledges the problem, but in resisting this move he fails to see *The Problem*. EM should not be a hypothesis or a theory or even a perspective, but the intellectual move that puts traditional intuitions about *cognition* into question, not just intuitions about boundaries. After the move is concluded, new genuine questions arise: how can we then make sense of a cognitive system as an agent, with a perspective, values, norms and even subjectivity? After tearing down the boundaries, in what sense can we recover an unprejudiced notion of an *individual* cogniser, a centre of activities and perspectives (and responsibilities), if at all? To believe that EM could eventually eliminate the mind altogether is to put too much trust on the intuitions about cognition that derive from traditional cognitive science. If they lead to dubious conclusions, maybe we should start looking for new ones before we stop asking questions that still make sense.

## 3. Enactive questions.

Enaction, Varela-style, starts with the question of the autonomy of the cogniser. The main blind spot in traditional cognitive science that this perspective addresses is the disappearance of any scientific, phenomenologically informed and useful notion of a cognitive system as an autonomous agent. In this, enaction is unsurprisingly similar to the theory of autopoiesis on which several of its key concepts are founded. This theory (Maturana and Varela, 1980) has as one of its main motivations to counteract the worrying evaporation of the organism from contemporary biology. Below the level of the organism (in biochemical processes and genetic networks) and above it (in evolutionary and ecosystemic dynamics) respectable science works unquestioned. At the level of the organism, it often sounds confusing, sometimes mystical (even in disciplines where this is a natural level of analysis such as embryology and ethology), and like anything confusing and mystical, the rational mind should eschew it from scientific discourse (and from reality altogether – joining the mind-sceptics, I have heard reputable biologists and bioengineers declaring that they do not believe that there is such a thing as life[1]). Psychology and cognitive science, with notable exceptions, have been in an identical situation with respect to the concept of a cognitive agent.

Already this question should alert us to the potential relevance of the (bio)enactive approach for addressing the problems I have put at the door of EM. But the situation is even more interesting. In recent developments, the enactive perspective has started to advance on the intimate connection between the concept of autonomy and sense-making, the normative engagement of a system with its world (Varela 1991, 1997, Weber & Varela, 2002, Thompson, 2007, Di Paolo, 2005, Di Paolo, Rohde, De Jaegher, forthcoming, De Jaegher & Di Paolo, 2007). The latter is nothing less than a strong candidate for a widely applicable, non-species-specific[2], non-bio-chauvinist *definition* of

---

[1] E.g., bio-chemist Andrew Ellington has recently made such a pronouncement at a keynote address at the 11th International Conference on Artificial Life in August, 2008, Winchester, UK.

[2] As it will become clear in the following, unlike the implicit human scope that EM reserves for what counts as cognitive, *non-species specific* for the enactive view means something more general than accepting as cognitive functionally equivalent implementations of mindful (human) capabilities in non-biological processes. It also means broadening the notion of cognition to encompass that of non-human species. EM is still human-chauvinist at the functional, if not implementational, level.

cognition. These, then, are the initial and current issues addressed by the life/mind branch of enactivism and they are precisely, as I said, the questions that EM fails to ask.

But ironically this very reliance on the theory of autopoiesis and on this crucial connection between living and cognitive processes has led Wheeler to declare enactivism as internalist. He interprets this reliance as a version of the statement: a cognitive system is a living system, and therefore cognition must happen inside the organismic boundary. What's going on?

## 4. From autopoiesis to sense-making

Wheeler's interpretation is not new, nor surprising given some of the statements that can be found in autopoietic literature. However, it is clear, particularly in the later work by Varela, that nothing like an internalist approach to mind is intended by the enactive programme. Cognition is an embodied engagement in which the world is brought forth by the coherent activity of a cogniser in its environment. Even in simple organisms, this engagement involves the structuring of the immediate milieu with the consequent building of regularities, which feed back into the organism itself. Even if we were to suggest that a cogniser must end at the organismic boundary (something that will not do), cognitive processes by necessity involve multiple feedback loops and organizing activity across that boundary. There is no place in the autopoietic literature where this is ever denied, and it is often said explicitly or implied. However, Wheeler is justified in taking autopoiesis to mean that a living system and a cognitive system are co-extensive. "Life = cognition" has been a famous motto that comes out of this literature (Boden, 1999, Bourgine and Stewart, 2004, Stewart, 1995). In addition, several statements in the primary writings seem to leave internalist clues: notably the emphasis of the physical boundary of the organism, when in fact what matters conceptually is always its organizational boundary, the insistence on the "non-intersection" of the constructive domain of constitution and the relational domain of interaction, when what is intended is "irreducibility", and the insistence on the closed organization (implying self-determination) of life and of higher forms of autonomous dynamics, such as the dynamics of the nervous system. These and other clues admit a non-internalist interpretation, which I think is the interpretation intended all along. However, they lend themselves to be used as effective building blocks for the story put forward by Wheeler. Instead of going case by case revisiting the main theory in order to justify non-internalist interpretations (which would require more than the available space), it will be better to stand on the side of the enactive approach and tell from this vantage point the story of exactly what is the role played by autopoiesis in this perspective and how it leads to a workable and unbiased definition of cognition.

### 4.1. Bare living.

The theory of autopoiesis is presented as a *biology of cognition*. The enactive approach uses some of the terminology of autopoiesis, in particular concepts such as autonomy and structural coupling. But it soon becomes clear that enaction is an ongoing development and that the theory of autopoiesis in itself should be critically assessed and, if necessary, re-interpreted or extended as a consequence of this development. It is a mistake to take the theory of autopoiesis as originally formulated as a finished theory (a trap that is easy to fall into because of the way the theory is presented in the primary literature). Several researchers have recently come to the realization that, as a theory of life and as theory of cognition, autopoiesis leaves many important questions unanswered (e.g., Moreno and Ruiz-Mirazo, 1999, Di Paolo, 2005, Thompson, 2007). In particular, several essential issues that could serve as a bridge between life and mind (like a proper grounding of teleology and agency) are given scant or null treatment in the primary literature, and questions about important biological phenomena such as illness or fatigue are not even raised.

In the original formulation the term autopoiesis refers to an organizational property of living systems: their physical self-production and self-distinction. This is proposed as the defining property of life. An observer should in principle be able to ascertain whether a system is autopoietic or not (for this, explicit operational criteria have been offered, e.g., by Fleischaker, 1988). Even though it refers to dynamical physical processes, the concept of autopoiesis does not make the time dimension explicit. The only way in which time enters in the theory is by the notion of *conservation* (i.e., invariance within a time interval). The system will always undergo structural changes while maintaining the condition of autopoiesis as long as autopoiesis is conserved. Otherwise, it disintegrates (a phrase that recurs in the primary literature). The reason why this tautology is of any interest is because the very identity of the system is given through the self-distinction introduced by autopoiesis. At this point the theory starts to rely on intuitions. Since it is not specified at all in the definition of an autopoietic system how this conservation of autopoiesis happens the theory is often complemented by appeals to metaphors. An autopoietic system is *like* a homeostatic system, whereby the homeostatic variable is its own organization. A close examination of the definition (Maturana & Varela, 1980, pp. 78-79) does not strictly lead to this conclusion. A homeostatic system connotes a notion of active monitoring and reaction to perturbations that challenge homeostatic variables which may sometimes go out of bounds. This may indeed be the case in certain autopoietic systems for some important variables, but it cannot be the case for its organization – for it to go out of bounds means that the system is dead. The definition of autopoiesis, moreover, does not rule out fortuitous, non-adaptive conservation of autopoiesis as in the case of a system that, without any compensatory mechanisms, just happens to be in a situation in which its self-production is unaffected. Maybe short-lived, maybe very fragile, but still (barely) autopoietic.

I have analysed this problem elsewhere (Di Paolo, 2005). It is precisely this issue, the non-entailment from the set-theoretic notion of conservation to the dynamical notion of homeostasis, that underlies many of the problems we find when we try to understand interactive and relational concepts like teleology, temporality, agency, sociality, and normativity (indeed most aspects of cognition). These have been the concern of enactivism since its inception in the later work of Varela. Part of this work can be read as attempts to reconcile the set-theoretic logic of autopoietic theory with the gradedness of concepts such as significance, norms, and values (e.g., Varela, 1997). His contention was that an autopoietic system, by the very fact that it is autopoietic, casts a veil of significance on its world. It *narrows down* the multidimensionality of a physical coupling into the distinction of whether the encounter is good or bad for autopoiesis (Weber & Varela, 2001)[3]. In other words, there is an intimate relation between being alive and being cognitive. The best interpretation at this point is not to say that there is a relation of equivalence between life and cognition but to say that a system that is living is also a system capable of cognitive engagements.

I believe this analysis can indeed be made of actual living systems but not from the plain fact that they are autopoietic. Varela's favourite example was bacteria swimming up a chemical gradient. It is their very organization (as well as their behaviour) that points us observers to the fact that sugar is significant to these organisms, while other chemicals are neutral or noxious. It also seems as if *more* of it is better than *less*. This is not a fact of the matter of the same kind as that describing a chemical reaction. It is a *relational* fact impossible to appreciate unless we have an organism present *for which* the

---

[3] I find the image of a reduction in the multiplicity of dimensions of a physical coupling somehow more accurate than that of an augmentation as implied in the phrase "a surplus of signification" used by Varela (1991, 1997). The implication that something is added by the organism to the encounter with the world downplays the richness of this encounter and the organism-making aspects of the world at the expense of the worldmaking aspects of the organism. As a consequence, the latter's role may be conceived unnecessarily as too active or cognitive. A narrowing down of the dimensionality of the encounter may also ultimately be an inaccurate metaphor, but it is more neutral with respect to the roles of organism and world.

effects of the chemical encounter on the processes of self-construction and self-distinction make sense as being good or bad.

The idea is nothing less than a route to the naturalization of values and norms, leading eventually to a naturalization of intentionality. However, the proposal cannot be reconciled with the primary literature on autopoiesis. In order for it to work, Varela must resort to notions such as "breakdowns" in autopoiesis, which may be major or minor (Varela, 1991). This runs against the conservation doctrine. Intuitive as such notions sound, they make no sense since autopoiesis is an all-or-nothing property. Autopoiesis does not come in degrees (Maturana and Varela, 1980, p. 94). Otherwise, all the talk about conservation would evaporate.

We simply cannot derive from the definition of autopoiesis that an autopoietic system will attempt to ameliorate a situation that leads otherwise to the future loss of autopoiesis (e.g., to swim *up* the sugar gradient). Improving the conditions of self-production is not part of the definition of autopoiesis, nor can it be deduced from this definition which simply states that the system self-produces. Hence, an autopoietic system that is able to operate according to potential future states cannot do so just because it is autopoietic. Given that cognitive systems often seem to be doing precisely this future-oriented regulation, it would seem that classical autopoiesis on its own and cognition cannot travel together from this point onwards. Through a different route, we arrive at something similar to the first part of Wheeler's conclusions: a *bare* autopoietic system cannot be a cognitive system and so it cannot be an extended cognitive system either.

### 4.2. Autopoiesis+.

It would be hasty to simply drop the conservation doctrine of autopoietic theory, even though it seems to be at the root of the problem. The set-theoretic analysis based on organizational properties (as opposed to only structural ones) is one of the strongest contributions of the theory. To get rid of it by softening the concept of autopoiesis (making it something relative and capable of partial breakdowns) amounts to reverting to a hazy view of living systems as being defined by a list of properties (growth, reproduction, responsiveness, etc.), the very view that autopoietic theory is trying to overcome. Moreover, such a move would do a biologically grounded theory of cognitive systems no favours since differences in cognitive performance or cognitive capability would be too easily married to metabolic differences and the resulting theoretical position would be weak. The cognitive domain would not be grounded on autopoiesis but reduced to it. This reduction would be unable to explain how meaning is often found in situations that do not directly affect metabolic production. In other words, the cognitive domain would lose its autonomy.

We must accept the terms of the problem and attempt a solution that will provide the required properties to approach our intuitive notions of cognition, and in so doing, possibly help us generate an operational definition of the term without appeals to the regulatory intervention of the inner as in the case of the parity principle. Such a property would help us make sense of Varela's notion of a breakdown in autopoiesis. This property should 1) come in degrees, 2) respond differentially to different situations according to their consequences for the organism, 3) sometimes malfunction, 4) obey the axiom of structural determinism and yet 5) allow the living system to alter its present operations with respect to non-actualized situations.

To arrive at this property the first thing that we must recognize are the observable differences that can be externally appreciated in autopoietic systems. We know when a sick organism has a few hours or days to live. Similarly, we can distinguish organisms living through risky or comfortable situations. Of course, none of these distinctions is perfect or absolute (the sick organism might recover, the risk of a situation might be illusory or its comfort belied by imminent but unperceivable danger). We as observers may normatively modulate *our* actions when we make such distinctions about other organisms. So it seems as if beneath the all-or-nothing viability condition given by the conservation doctrine lies a space of graded and qualitative structural differences from which norms may be actualized in a cognitive act of observation, e.g., it is generally better to avoid risky situations than to seek them, health is better than sickness, and so on. What our story requires is not for us observers but for the organism itself to be able to generate such norms while it is still alive and to regulate its operations accordingly (to *evaluate* the situation) within the space of structural options that corresponds to the conservation of autopoietic organization. This capability describes the property of *adaptivity*.

Autopoietic systems exist far from equilibrium and must tolerate the natural entropic trends by remaining energetically and materially open. In other words, they are robust in that they can sustain a certain range of perturbations as well as a certain range of internal structural changes without losing their autopoiesis. These limits are defined by the organization and current state of the system and are here referred to as its *viability set* (we can often measure or estimate some aspects of this set, for instance, in variables that must be kept within certain bounds, like blood temperature in mammals). If the trajectory of states approaches the limits of viability and moves outside them, the system dies. Adaptivity is then a special manner of being tolerant to challenges by actively monitoring perturbations and compensating for their tendencies. Adaptivity is defined as (after Di Paolo, 2005, p. 438):

> a system's capacity, in some circumstances, to regulate its states and its relation to the environment with the result that, if the states are sufficiently close to the limits of its viability,
>
> 1. tendencies are distinguished and acted upon depending on whether the states will approach or recede from these proximal limits and, as a consequence,
>
> 2. tendencies that approach these limits are moved closer to or transformed into tendencies that do not approach them and so future states are prevented from reaching these limits with an outward velocity.

An adaptive autopoietic system is able to operate differentially in (at least some) situations that, were they left to develop without any change, would lead to loss of autopoiesis. This property is perfectly operational but not implied in the definition of autopoiesis. It allows us to recover a homeostatic interpretation. A *breakdown*, for the system, is simply the severity of a negative tendency (a tendency of states to approach the proximal limits of viability) distinguished and measured by the time constraints it imposes and the amount of regulative resources that it demands to compensate for it. A breakdown will typically, but not exclusively, be the result of external perturbations, and in addition to responding to them, adaptivity may allow organisms the possibility of avoiding some risky situations and seeking preferable ones[4].

Only of the subset of autopoietic systems that are not just robust but also adaptive can we say that they posses operational mechanisms to potentially distinguish the different virtual implications of otherwise equally viable paths of encounters with the environment. This differential operation is

---

[4] It should be noted that this definition is meant to be quite general. Systems that are adaptive in this sense can come in many variations (adaptive to only a few perturbations, but not to others, adaptive only in reaction to a harmful event or able to foresee and prevent its occurrence, adaptive in a highly reliable manner or adaptive in a rough statistical sense, etc.). The definition also implies that in practice observers can agree on a metric or at least sufficient topological properties of the viability set as well as on what counts as the appropriate granularity and timescale of observation and that this agreement reflects well the notion of risk or danger for the organism.

called *sense-making* (Weber & Varela, 2002, Thompson, 2007, Di Paolo, 2005). If, as proposed, sense-making requires the acquisition of "a valence which is dual at its basis: attraction or rejection, approach or escape" (Weber & Varela, 2002, p. 117), a system engaged in sense-making requires, apart from the norm given by self-construction, access to how it currently stands against the all-or-nothing barrier given by that norm. In order to have such access while operating as a consequence of contemporaneous states, an autopoietic system must be able to recognize in those states, and only in them, the virtual tendencies that relate it as a whole to the potential loss of its own viability and, in addition, it must be able to act appropriately on those tendencies. This is the basis for a normative regulation of organismic operation, including how the organism regulates as a whole unity its interactions with the environment. Such normative engagement with the virtual consequences of current tendencies is the hallmark of cognition (though as we shall see, not all norms refer back to the logic of metabolism).

Adaptivity not only provides a conceptual tool that paves the way from autopoiesis to cognition, it helps us thematize several other biological phenomena while remaining within the perspective of autopoietic theory. These are more extensively discussed in (Di Paolo, 2005). Crucially for the present discussion, when adaptive mechanisms operate across the physical boundary of an organism so as to regulate its coupling with the environment, we move from *structural coupling* (essentially a symmetrical concept whereby system and environment influence each other without loss of viability) to *behaviour* (an asymmetrical concept where the organism originates the regulation of structural coupling). This regulation of interactions allows us to define certain adaptive autopoietic systems as *agents*. The concept of agency is asymmetrical because it refers to a modulation of coupling, which is already a loop of mutual constraining between processes pertaining to the organism and processes in the environment. Modulation therefore implies a second, higher-order loop following the norms of the agent's adaptivity, (Figure 1). This second-order loop is rarely made explicit (though it already appears as a central idea of Ashby's theory of adaptive behaviour, Ashby, 1960). The resulting distinction between an engagement and the normative regulation of this engagement is not present in the EM literature. Thanks to it, we may now define *cognition* as *sense-making in the interactive domain*. In this way, we capture most (perhaps all) of our important intuitions about the term: it is normative, asymmetric (there is a centre of cognitive activity), relational, oriented and extended in time, capable of mistakes, and it implies a self-constituted identity to which norms refer.

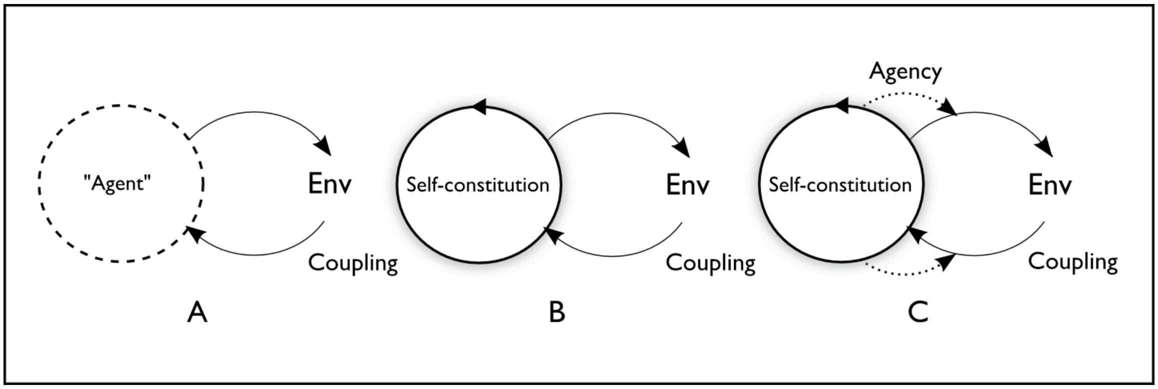

Figure 1. Views on agency. Case A indicates the typical usage of the term "agent" in robotics and cognitive modelling, i.e., a system defined by convention in coupling with the environment (Env) and this coupling is symmetrical; in B the system is replaced by a self-sustaining process of identity generation, such as bare autopoiesis which is not yet an agent; C illustrates the idea of agency as the modulation of the coupling between the autonomous entity and its environment. (Notice that the direction of the regulatory arrows indicate a modulation of both "sensor" and "effector" aspects of the coupling. The necessary feedback for this regulation is not included in the diagram and may be assumed to result from the monitoring of the primary loop).

### 4.3. Rooted in life, cognition moves on its own.

Even though it is possible to open up the theory of autopoiesis by introducing the concept of adaptivity, there are still problems with grounding cognition solely in adaptive autopoiesis (e.g., What about norms and values that guide behaviour and are underdetermined by metabolism? Isn't the possibility of adaptive dysfunction an indicator of a merely contingent connection between life and cognition?). Addressing these problems implies adopting a wider perspective, one that permits us to thematize the irreducibility of the cognitive domain to its enabling metabolic substrate (and dispel once and for all the interpretation of co-extensiveness between the living and the cognitive).

The origins of sense-making are connected with the regulatory capacities of an organism, but more generally with the presence of a process of identity generation. This is the idea of autonomy that we adapt from (Varela, 1979) to include a requirement of *precariousness*. Accordingly, an *autonomous system* is defined as a system composed of several processes that actively generate and sustain an identity under precarious circumstances. In this context, to generate an identity is to possess the property of *operational closure*. This is the property that among the conditions affecting the operation of any constituent process in the system there will always be one or more processes that also belong to the system. And, in addition, every process in the system is a condition for at least one other constituent process, thus forming a network. In other words, there are no processes that are not conditioned by other processes in the network – which does not mean, of course, that external processes cannot also influence the constituent processes, only that such processes are not part of the operationally-closed network as they do not depend on the constituent processes. Similarly, there may be processes that are influenced by constituent processes but do not themselves condition any of them and are therefore not part of the operationally-closed network. In their mutual dependence, the network of processes closes upon itself and defines a unity that regenerates itself (in the space where these processes occur). Precarious circumstances are those in which isolated constituent processes will tend to run down or extinguish in the absence of the organization of the system in an otherwise equivalent physical situation. In other words, individual constituent processes are not simply conditioned (e.g., modulated, adjusted, modified, or coupled to other processes) but they also depend for their continuation on the organizational network they sustain; they are enabled by it and would not be able to run isolated.

It is important to note that the concept of precariousness plays a crucial role in this approach. The added requirement of precariousness is essential in order to make the idea of operational closure non-trivial and thus turn it into an effective tool for generating explanations that capture the phenomenology of life and mind. Self-sustaining patterns of activation in an artificial neural network or in the toroidal cellular automaton Bittorio (Varela, et al, 1991) might qualify under the proper interpretation as operationally closed. However, if we take as relevant constitutive processes the state transformation rules in these dynamical systems, these rules are not themselves enabled by the organization of the self-sustaining patterns but are independent of them. If and when an operationally closed system is established out of processes that not only sustain themselves as a network, but additionally, would not be able to sustain themselves individually unless they were organized as such a network, we are in the presence of a precarious operationally closed system. Fundamentally, in such a system component processes will not be present or durable enough, unless the conditions that are sustained by the rest of the network are in place. Such an operationally closed system not only maintains itself; it enables itself. It produces its own preconditions. As a consequence the processes that compose such an autonomous system blur the distinction between operator and that which is operated upon. Each process must be prevented from decay by the activity of other processes whose operation is affected by the specific way in which they themselves are prevented from decay. In having to act against the natural tendencies of its own material support, a precarious autonomous system must produce a proper self-positing response to these tendencies.

This response does not just happen to require material resources (as is the case of all far-from-equilibrium systems), but its very form depends on the specific material precariousness it is fighting against. A precarious autonomous system, at whatever level, intervenes in its own substrate in order to sustain a form which is made out of components that paradoxically *provide the very tendencies towards the dissolution of the same form*. This persistence through (double) self-negation is at odds with functionalism – understood generally as attempts to provide a substrate-*independent* account of the operation of a system. Precariousness does not refer to a positive material property that could be captured functionally, but to the impermanence of any relevant positive property of the substrate[5].

With the notion of autonomy in place, a first point that needs to be addressed is the logic that links an autonomous process of precarious identity generation and the normative, teleological, value-laden relation between this identity and its medium (i.e., sense-making). Adaptive autopoiesis is but one instance (perhaps the most fundamental) where we witness this link at work. Hans Jonas has explored precisely this logic (Jonas, 1966). The fact that metabolism sustains a dynamic form of identity (not coinciding with its material constitution at any given time except at the time of death) allows an organism to *become* free. This freedom is expressed in the capability of the organism to engage with its medium in terms of the significance of a situation, thus contributing to its continuing dynamical autonomy and even opening up the possibility of novel value-making. However, this freedom is allowed by very strict and specific material needs. It is a *needful freedom*. Rather than being paradoxical, this concept of freedom avoids the problems posed by determinism (and indeed by randomness) by operating on the relation of *mediation* between the self-sustaining identity and the "target" of its cognitive engagements. In that sense, the mode of realisation of an autonomous process of identity-generation (like autopoiesis) establishishes what sort of access this identity has to the norms that describe its different modes of viability. This access may be less or more mediated (the difference, say, between reacting with aversion to contact with hot surface vs. planning our movements so as to avoid touching it, both of which are examples of adaptive regulation). Jonas' contention is that in the history of life and mind novel forms of increasingly mediated engagements have appeared allowing for more freedom at the cost of more precariousness.

Animals are a clear, but not the only example of such transitions. A new order of values is found in animality with the advent of motility and the co-emergence of perception, action and emotion. By putting a distance and a lapse between the tensions of need and the consummation of satisfaction, the temporality of adaptivity (Di Paolo, 2005) is spatialized. Animals can appreciate right now the danger that is impinging on them from a distance. This is the origin of a special relation with the world, that of perception and action, which is charged with internal significance, and hence with the development of an emotional dimension (what might have been an inner life of need and satisfaction now becomes rich in possibilities such as fear, desire, apprehension, distension, tiredness, curiosity, etc.). But this comes at a cost of more severe energetic demands (allowing the necessary fast and continuing movement across varying environmental conditions without replenishment for long periods) and novel forms of risk. For instance, several species of insects, like the water boatman, are able to breathe underwater by trapping air bubbles (plastrons) using tiny hairs in the abdomen. The bubbles even refill with oxygen due to the differences in partial pressure provoked by respiration and potentially can work indefinitely, (see e.g., Thorpe 1950)[6]. They provide access to longer periods

---

[5] Life would not be better off without precarious conditions; it simply would not be life at all. It would be indifferent permanence. In contrast, precariousness presents us with a view of life as inherently restless more fitting to our experience of life; a sort of "frustrated suicide", never fully safe and constantly buying time for itself. Notably, it is here that the enactive project most clearly departs from any form of functionalism including the extended functionalism proposed for the EM: by its appeal to an aspect of materiality which is inescapable and *nevertheless still required* for a theory of life and mind that can account for autonomy and normativity.

[6] Thanks to Susan Oyama for providing this example.

underwater thanks to a mediated regulation of environmental coupling (which is nevertheless potentially riskier than normal breathing). The mediation in cases like this is so intimately connected with vital functions that the living system itself might be called extended. The issue at play in such reliable and conserved forms of mediation is, in each case, the question of the identity of such extended systems.

Jonas recognizes other such transitions in modes of mediation in the history of life and mind, such as for instance those afforded by a complex visual system or the capacity to make images. It is clear that no intrinsic gain is implied at the metabolic level by expanding the realm of freedom at the cost of increased precariousness. However, he points out that "the survival standard is inadequate for an evaluation of life" (Jonas, 1966, p 106). He goes on:

> "It is one of the paradoxes of life that it employs means which modify the end and themselves become part of it. The feeling animal strives to preserve itself as a feeling, not just a metabolizing entity, i.e., it strives to continue the very activity of feeling: the perceiving animal strives to preserve itself as a perceiving entity – and so on. Without these faculties there would be much less to preserve, and this *less* of what is to be preserved is the same as the *less* wherewith it is preserved" (ibid).

Effectively, such transitions in mediacy inaugurate a domain that feeds back on itself; they imply a *new form of life*. Not just in a metaphorical sense, but in the strict sense of a novel process of identity generation underdetermined by metabolism.

But how is this possible? Can we make sense of this in terms of autopoiesis? The problem of how to connect the constructive and the interactive aspects of living organization is already inherent in the phrasing of autopoietic theory. This difficulty is hard to appreciate (let alone resolve) from within the terms proposed by the theory. The problem is the impossibility of transversing the operational and the relational domains. The first pertains to the functioning of the autopoietic network so that it constitutes a unity (a composite system), the second to the relations that such a unity enters into in its structural coupling with the environment (see, e.g., Maturana, 2002). For all the logical accountancy that this separation into so-called "non-intersecting domains" affords, the theory has not dwelled on the problem that this separation brings, i.e. a systemic analogue to mind-body dualism. In effect, the theory of autopoiesis says nothing about how relational interactions and internal compensations are coordinated. They just happen to be or else there is no autopoietic system. So, in this specific interpretation (irreducibility as non-intersection), autopoietic theory is strictly Cartesian in a way that Jonas and enactivism (and even Maturana and Varela themselves!) are trying to avoid. The Malebranchean solution – i.e., that God takes care of the coordination between non-intersecting substances, – here takes the form of appeals to evolution, appeals that are pregnant in the phrase "otherwise it disintegrates". Evolution takes care of sieving out those unhappy organisms for which the two domains are uncoordinated[7].

---

[7] The intention behind the distinction between domains should be clear: to prevent any attempt at reducing phenomena across these domains. Given that one domain is established by the presence of a whole unity and its relations to its environment, there are good systemic reasons to distinguish those relations from the constitutive processes that give rise to the unity. To reduce phenomena across these domains is to ignore the alternative perspectival positions one may take on a given system: that of a whole unity in relation and that of a multiplicity of components that make up the unity. Such confusion would lead us to search for the speed of the car inside its engine (or to describe our memories as represented in neural patterns). This is a strong point that should be preserved. The damage, however, is done by the term "non-intersecting". This connotes strict separability whereas in fact, we can certainly maintain a relation of non-reducibility without isolating the phenomena between domains. In this way, it is indeed possible for explanations in domain A to depend on phenomena in domain B, but not exclusively so; a powerful engine conditions the possibilities for explaining the speed of the car even if we cannot deduce the latter exclusively from the former. Where we must be careful

The major problem that is apparent with this solution is that the relation between the two domains is purely contingent. It happens to be like this because it has helped the system survive. This falls short of grounding the cognitive in the systemic. For, without denial of the role of evolution, it is clear that grounding mind in life requires establishing the necessary links between phenomena in these two domains. What an organism is and what it does should not be properties external to each other. Otherwise, the project fails. By contrast, the Jonasian solution is that a transition to a sustained new form of value-making (such as in animality or image-making) regenerates the very organizational conditions that made the transition possible. It either changes the form of identity generation that sustains the new interactive domain, or indeed it establishes a new form of autonomous identity (the *feeling* animal, the *perceiving* animal, etc.).

Despite the problems just highlighted with bare autopoietic theory, the idea of other forms of autonomy in terms of operationally-closed dynamics apart from autopoiesis is indeed an acceptable possibility. The theory highlights the operational closure of the nervous system (Maturana and Varela, 1980, p. 127) and the immune network (Varela, 1991) and Varela has suggested that other domains may possess similar forms of autonomy, albeit not in terms of relations of material production. Such could be the case of sensorimotor loops, conversations and social interactions in general (Varela, 1979, 1991, 1997, De Jaegher & Di Paolo, 2007). However, what is left unsaid is the kind of relations that such identities may bear to one another. The transformative relations between constructive and interactive aspects of autonomy leading in themselves to a novel form of identity cannot be directly addressed by autopoietic theory. This is simply because a logical barrier is put between the two domains and because an emphasis on conservation of autopoiesis obscures the possibility of a *structural becoming* of novel forms of organization encompassing both constructive and interactive aspects of the living.

We can now understand why the transitions in mediacy described by Jonas (and several others not made explicit by him) have an irrevocable character. They are authentic births of new lifeforms. These new lifeforms may relate to the metabolic substrate and other lifeforms in a variety of ways, calling for a veritable topology of processes of identity generation (intersecting, embedded, hierarchical, shared, etc.). It is also an open possibility that the dependence on a form of life so much modifies the basic autonomy of metabolism that the higher identity essentially intervenes in the very condition of operational closure of autopoiesis, Di Paolo, forthcoming).

From a systemic perspective, the relation between different self-sustaining processes enabled by a substrate of autopoiesis need not be one of perfect harmony. On the contrary, the inherent regulative tendencies of sophisticated processes of identity generation are likely to sometimes enter into conflict even with basic metabolic values. I and others have proposed that, in their plasticity, habits should be seen as such autonomous structures (encompassing partial aspects of the nervous systems, physiological and structural systems of the body and patterns of behaviour and processes in the environment), (Barandiaran, 2008, Di Paolo, 2003, Di Paolo, 2005). And habits, as we know, can be "bad". The question is that as self-sustaining structures, they are never bad for themselves, but for some other identity (typically, in the case of humans, a combination of the metabolic and socio-linguistic self).

This is where the (bio)enactivist story may speak directly to the EM theorist. New forms of life tend to conserve in modified form the enabling levels that give rise to them. The best image is that of embedded systems whose identities participate in their mutual conservation, but may potentially also enter into conflict. The dependence and mutually enabling constraints between different forms of

---

is in the form that such dependence takes since any relation across domains will always be a relation of modulation or constraint, and not of determination.

identity (say between metabolism and habits brought about by the possibility of agency) may result in an integrated system. This system is not coextensive with the basic level identity. This life/mind system (Di Paolo, forthcoming) conserves life (autopoiesis) but is different from it and may indeed involve the incorporation of relations of mediation (agential involvements that recur and self-sustain) into its own constitution. This is in the spirit of EM but is, at the same time, more general and more concrete. More general because the new identity is not necessarily an extension of the original identity (e.g., autopoiesis) that existed before the addition of self-sustaining mediation. It is simply a different system – in some cases, like the underwater breathing of insects, or the blind man with the cane, the image of an extended system is highly suitable, in others, such as a mammal that learns the habit of climbing trees and navigating though the forest canopy, the novel form of life is not the result of a systemic extension in the EM sense. The possibility is also more concrete because it calls us to distinguish specifically between regulated engagements that are part of the constitution of a new identity from those regulated exchanges that are not. This can be done in principle through the relation of operational closure under precarious conditions that must verify in order to generate an identity of any order.

This is yet another distinction to be added to the EM toolbox, a cognitive transformation need not be merely a spatial extension (not even, metaphorically, always an augmentation) of the original system, it only needs to be a self-conserving change. However, the distinction in principle might still be problematic in practice. In humans, habits form organized bundles (Barandarian, 2008, Barandarian & Di Paolo, 2008) whose organization into closed networks still requires better theoretical elucidation. There is a sense in which it is true to affirm that the agent that I am when I'm swimming is different from the agent that I am when I'm taking photographs. At this point, social interactions may intervene in stabilizing one identity with several capacities as opposed to several regional identities. Otherwise, systemically speaking, the continuity of a single cognitive identity across different forms of agential engagement would be difficult to establish, since it would not suffice to provide links to the underlying metabolic identity which as we have seen can be involved in several closed self-sustaining networks. It is potentially the action of other cognitive selves and the general background of significance of human activities that helps sustain a socio-linguistic identity as that of a body plus a capacity for multiple engagements associated with it. The human cognitive self is thus co-determined with the other (see also Thompson, 2001, De Jaegher & Di Paolo, 2007).

## 5. Back to EM

In his argument, Wheeler establishes the relation of autopoiesis to sense-making as one of necessity, not sufficiency between the former and the latter. This is not incorrect but in the view presented here the statement requires a clarification. It is in fact the identity-generation aspect of autopoiesis (its autonomy) that grounds the normativity upon which (with the addition of adaptivity) sense-making is possible. This is important, since other forms of autonomy, not necessarily involving processes of molecular production and transformation like metabolism, may sustain a similar normative regulation of coupling (e.g., patterns in neuronal ensembles that sustain certain behaviours and are only prevented from decaying by the plastic neuronal effects provoked by the enaction of such behaviours). The Jonasian story leads us into a way of resolving the complex entanglement between relation and constitution. Cognition is always relational, but it is certainly possible for relational processes to be involved in closed networks that constitute a novel identity. For this, we need a prior identity to which relations originally refer. In some cases, the new identities may deserve the name of extended systems when their constitution relies on relational processes of mediation and normative regulation between an underlying system and its environment. The assumed sufficiency of cognition for life, upon which Wheeler builds his final argument (by establishing that being a cognitive system is sufficient for being a living system) is then put into question, since autonomy may occur in systems that are not themselves autopoietic (a cognitive system in this case would be enabled by metabolism *and* certain reliable forms of coupling forming a closed organization). Under this view, the tension identified by Wheeler is resolved.

Even as a rushed tour of current enactive thinking on the relations between life and mind, I hope that the foregoing paragraphs provide enough evidence to conclude that the proposed grounding of cognition in life is quite different from a simple statement of co-extensiveness of the living and the cognitive. Nor is it at any point possible to infer an internalist view of cognition in the enactive story. Indeed, neither an externalist one! Cognition is sense-making in interaction: the regulation of coupling with respect to norms established by the self-constituted identity that gives rise to such regulation in order to conserve itself. This identity may be that of the living organism, but also other identities based on other forms of operationally closed networks of processes, such as socio-linguistic selves, organized bundles of habits, etc. Some of these identities are already constituted by processes that extend beyond the skull. But in any case, cognition is always a process that occurs in a relational domain. Unlike many other such processes (e.g. getting wet in the rain) its cognitive character is given normatively and asymmetrically by the self-constituted identity that seeks to preserve its mode of life in such engagements. As relational in this strict sense, *cognition has no location*. It simply makes no sense to point to chunks of matter and space and speak of containment within a cognitive system. Inspect a baby all you want and you'll never find out whether she's a twin[8].

Does any of this go against the EM? Strictly speaking, yes. The enactive approach does provide a workable and non-species-specific definition of cognition, it does not fall into paradoxical situations of breaking boundaries while re-introducing them through the back door as regulatory principles, and it thematizes norms, identities, mediation and agency in a way that, in their reliance on the precariousness of materiality, are incompatible with the extended functionalism of EM (as we have seen, precariousness cannot be accounted for functionally, and this is the reason why the virtuality and temporality of sense-making are also lost to functional accounts in general and EM in particular). Enactivism thus comes close to Malafouris' (2004) concept of materiality as intrinsically defining the experience of a cognitive engagement. Wheeler significantly dubs this concept *vital materiality* (2008).

However, I think that none of what I would take as the really interesting lessons of EM are lost in the enactive account. Cognition is not skull-bound and different extra-bodily processes are inevitably participants in cognitive engagements. Indeed, it is now possible to introduce further distinctions and question whether a given part of the world simply contributes to a cognitive coupling or to the regulation of this coupling, i.e., whether it is part of the cognitive process or more specifically part of the agency at work in that process.

We can list some of the contributions of the enactive approach to the EM story.

1. A workable, unbiased definition of cognition.

2. A distinction between relational and constitutional domains.

3. The notion of autonomy as a self-sustaining identity under precarious conditions.

---

[8] And yet, we must repeat that the relational and operational domains, while irreducible to each other, are still able to condition and constraint each other. The twin case, strictly valid for identical twins, reminds us of a rare but perfectly consistent illustration. It may happen that dizygotic twins show chimerism in their blood. During gestation blood cells from the opposite twin can be acquired *in utero* and they can initiate a successful, but genetically different lineage that can persist throughout a lifetime (Booth, et al, 1957). While not a definitive proof (maternal blood cells could also be acquired in this way, or indeed the embryo's blood cells be acquired by the mother in the case of microchimerism), blood chimerism in an individual would lead us to suspect her relational condition as a (non-identical) twin. A history of relations can indeed leave a mark in the constitution of a system.

4. The grounding of sense-making and normativity in the process of adaptive identity generation.

5. The distinction between coupling and asymmetric regulation of coupling leading to a definition of agency and behaviour.

6. A workable concept of mediation (as the marker to follow in locating complex forms of identity generation).

7. The notion of a topology of regional identities.

8. A hint towards the role of social interaction and participatory sense-making in estabilizing the socio-linguistic cognitive self.

These notions open up a potentially accessible field of problems to be solved. For instance, how to distinguish, using the Heideggerian terms, between the serviceability of a tool and the capacity of an organ (a constitutive part of an identity) (Heidegger, 1983/2001)? How to explain the experience of transparency in skilful coping (see also Thompson and Stapleton, this issue)? What is the developmental course through which a novel regional identity is produced? In what specific ways might artificial, non-living systems, be/become cognitive? And so on.

The more interesting and forward-looking themes introduced by the EM approach, and towards which enactivism must still develop, include the problem space of technical individuation and technological networks that bootstrap the generation of cognitive identities. This is the notion of a medium structured by the activity of cognisers past and present, which may itself be an active structuring factor in the formation and re-generation of similar patterns of cognitive engagement in agents that previously were not capable of performing them. Such "living" media show aspects of operational closure at collective levels without necessarily showing individuality. These include patterns of parental scaffolding, social reproduction structures like Bourdieu's *habitus* (1990), epistemic form of niche-construction, stigmergy and indeed – the links between life, mind and society flow both ways – "down to" the processes that occur in the extra-cellular matrix of multicellular organisms, or in biofilms in bacterial collectives. Such cases put in evidence that the richness of environments as active media has so far been underplayed in the current enactive story. EM has sought to thematize this richness and the enactivist should listen.

## Acknowledgements.

Thanks to Julian Kiverstein, Michael Wheeler, Hanne De Jaegher, Matthew Egbert, Arantza Etxeberria, Giovanna Colombetti, Tom Froese, and Evan Thompson for their helpful comments on this paper and to participants of the *Life and Mind Seminars* (http://lifeandmind.wordpress.com) where some of the ideas expressed here have been discussed.

## References.

Ashby, W. R. (1960). *Design for a Brain: The Origin of Adaptive Behaviour* (Second edition). London: Chapman and Hall.

Barandarian, X., (2008) Mental life: A naturalized approach to the autonomy of cognitive agents. Unpublished PhD Thesis, University of the Basque Country, Spain.

Barandiaran, X. and Di Paolo, E. A. (2008). Artificial mental life (Abstract). In S. Bullock, J. Noble, R. A. Watson, and M. A. Bedau (Eds.) Proceedings of the 11th International Conference on Artificial Life, Alife XI, Winchester, UK, MIT Press, Cambridge, MA.

Boden, M. A. (1999). Is metabolism necessary? *Brit. J. Phil. Sci.* 50: 231–248.

Booth, P. B., Plaut, G., James, J. D., Ikin, E. W., Moores, P., Sanger, R. and Race, R. R. (1957). Blood chimerism in a pair of twins. *Br Med J.* 1(5033): 1456–1458.

Bourdieu, P. (1990). *The Logic of Practice.* Cambridge: Polity Press.

Bourgine, P. and Stewart, J. (2004). Autopoiesis and cognition. *Artificial Life* 10: 327–346.

Clark A. (2007). Curing cognitive hiccups: A defense of the extended mind. *Journal of Philosophy,* 104, 163-192.

Clark, A. (forthcoming). Memento's revenge: The extended mind, extended In Menary R. (Ed) *The Extended Mind.* Aldershot: Ashgate.

Clark, A. & Chalmers, D., (1998). The extended mind. *Analysis* 58 (1), 7-19.

De Jaegher, H. and Di Paolo, E. A. (2007). Participatory sense-making: An enactive approach to social cognition, *Phenomenology and the Cognitive Sciences*, 6(4), 485 - 507.

Di Paolo, E. A., (2003). Organismically-inspired robotics: Homeostatic adaptation and natural teleology beyond the closed sensorimotor loop. In Murase, K., & Asakura, T. (Eds.), *Dynamical Systems Approach to Embodiment and Sociality*, pp. 19–42. Advanced Knowledge International, Adelaide.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4, 97-125.

Di Paolo, E. A. (Forthcoming) Overcoming autopoiesis: An enactive detour on the way from life to society. In R. Magalhaes and R. Sanchez, *Autopoiesis in Organizations and Information Systems,* Elsevier.

Di Paolo, E., Rohde, M. and De Jaegher, H. (2008). Horizons for the enactive mind: Values, social interaction, and play. In *Enaction: Towards a new paradigm for cognitive science.* Stewart, J., Gapenne, O. and Di Paolo, E. (eds) Cambridge, MA: MIT Press, in press.

Fleischaker, G. R. (1988). Autopoiesis: the status of its system logic. *BioSystems*, 22(1), 37 – 49.

Heidegger, M. (2001). *The Fundamental Concepts of Metaphysics: World, Finitude, Solitude.* Translated by William McNeill, Nicholas Walker, Indiana University Press.

Jonas, H. (1966). *The phenomenon of life: Towards a philosophical biology*. Evanston, IL: Northwestern University Press.

Malafouris, L., (2004). The cognitive basis of material engagement: where brain, body and culture conflate, in *Rethinking Materiality: the Engagement of Mind with the Material World*, eds. E. DeMarrais, C. Gosden & C. Renfrew. Cambridge: McDonald Institute for Archaeological Research, 53-61.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: A history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, 9, 5–34.

Maturana, H. and Varela, F. J. (1980). Autopoiesis and cognition: The realization of the living. Dordrecht, Holland: D. Reidel Publishing.

Moreno, A. and Ruiz-Mirazo, K. (1999). Metabolism and the problem of its universalization. *BioSystems* 49: 45–61.

Rupert, R., (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy* 101: 8, 389-428.

Stewart, J., (1995). Cognition = life: Implications for higher-level cognition. *Behavioural Processes*, 35(1-3), 311 – 326.

Thompson, E. (2001). Empathy and Consciousness. *Journal of Consciousness Studies*, 8, pp. 1-32.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind.* Harvard University Press.

Thompson, E. and Stapleton, M. (this issue). Making sense of sense-making: Reflections on enactive and extended mind theories.

Thorpe, W. H. (1950). Plastron respiration in aquatic insects. *Biol. Rev.* 25, 344–390.

Varela, F. J. (1979). *Principles of biological autonomy*. New York: Elsevier, North Holland.

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In: A. I. Tauber (ed.), *Organism and the origin of the self*, pp. 79–107. Netherlands: Kluwer Academic Publishers.

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition* 34, 72–87.

Varela, F. J., Thompson, E. and Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.

Weber, A. and 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. (2008). Mind, things and materiality. In Renfrew C. and Malafouris L. (eds.), *The Cognitive Life of Things: Recasting the Boundaries of the Mind*, McDonald Institute for Archaeological Research Publications, Cambridge.

Wheeler, M. (forthcoming). Extended X: Recarving the Biological and Cognitive Joints of Nature. Book manuscript. http://www.philosophy.stir.ac.uk/staff/m-wheeler/ExtendedX.php, last accessed 22 September, 2008.