CHAPTER 4

THE ENACTIVE CONCEPTION OF LIFE

EZEQUIEL A. DI PAOLO

A PROJECT OF THE WORLD

A widely known quote by Maurice Merleau-Ponty encapsulates the dialectical relation between subject and world in a succinct formula:

> The world is inseparable from the subject, but from a subject who is nothing but a project of the world; and the subject is inseparable from the world, but from a world that it itself projects. (Merleau-Ponty 1945, p. 454)

Meant to overcome the opposing pulls of idealism and empiricism, this statement contains in its formulation—if not in its meaning—a pregnant mystery.

In the context of this phrase, we are told that while the perceptual world may be posited by consciousness, this is only possible if there is underlying this consciousness an already oriented body engaged in transactions with the world. The perceiving subject is not an absolute world anchor, since all the existence bestowed on this world as a totality of meaning through sense-giving activity is inextricably entangled in the ways the world gives itself to the subject as a person, but also as an animal, as a living organism, and as a complex stream of material flows and potentialities. Another way to put it is that the subject is *in* the world but also *of* the world.

It is indeed a very enactive thing to say that the subject lays down the path in walking, that is, that the frames of signification are given by sense-making activity itself, which is by nature transactional and constrained within material and historical possibilities.

So far, so good. But what is the mystery in Merleau-Ponty's resounding formula? I find it in the idea of the subject as being a *project* of the world. Taken literally, it is a dissonant turn of speech as only those things are projects that are some*body*'s project.

And the world is not an entity that projects anything—if we decide to stand firm on non-teleological ground and avoid seeing nature as a whole as directed toward ends.

In what ways could then this be more than figurative language metaphorically referring to the transactional, materially constrained aspects of sense-making already mentioned? I will suggest in what follows that, as is often the case, a mysterious, yet beautiful, formulation invites a deeper truth. In the context of embodied perspectives on cognitive phenomena, this truth has been the concern of enactivist researchers. For them it has become clear that to ask questions about how the mind works is at the same time to ask questions about what is it about certain entities that they can be minds at all, and how can such entities emerge in a natural world. These two questions, which might be divorced in other areas of inquiry, are for the enactive perspective one single question with different facets. Hence the insistence on the part of some enactive thinkers on the need to understand life and mind as part of a continuity.

Differently put, I am talking about the difficult question: what is a body? This question, not always put in these explicit terms, is the platform on which enactive theory[1] is raised. It is, in my opinion, what differentiates the enactive approach from all other so-called embodied approaches: the thematization of bodies as a prerequisite for understanding *anything* about minds.[2] This is not a line of theorizing that emerges

---

[1] By enactive theory I refer here to the application of the enactive approach to specific scientific problems in psychology, neuroscience, cognitive science, AI, etc. Examples that explicitly use the label enactive include the development of a dynamical systems interpretation of sensorimotor contingencies theory (Buhrmann et al. 2013; Di Paolo et al. 2014; Di Paolo et al. 2017), nonrepresentational accounts of the phenomenology of the sense of agency (Buhrmann and Di Paolo 2017), neurophysiological models of multi-joint movements (Buhrmann and Di Paolo 2014), clarification of explanatory roles of social interaction in psychology (De Jaegher et al. 2010), hypotheses on social brain function (Di Paolo and De Jaegher 2012), hypotheses on sociocognitive development (e.g., Gallagher 2015), integrative, person-based approaches to autism (De Jaegher 2013), accounts of intentionality, action, and free will (Gallagher 2017), models of metabolism-based bacterial chemotaxis (Egbert et al. 2010), hypotheses on the interactive factors affecting imitation (Froese et al. 2012), organism-based theories of color vision (Thompson et al. 1992), general perspectives on brain function (Fuchs 2011, 2017; Gallagher et al. 2013), accounts of synesthesia (Froese 2014), accounts of neurobiological and embodied factors in prehistoric art and material culture (Froese et al. 2013; Malafouris 2007, 2013), and others.

[2] Unfortunately, the term "embodied" has become one of the most abused keywords in cognitive science. It is not necessary to produce a sophisticated critique of this term in order to see that whatever legitimate meaning it used to have has now been relentlessly diluted thanks to its adoption by brain-centered, individualistic, representational theories that are veiled versions of computationalism. Tenuous conceptual connections with the body do not make a classically disembodied approach any less disembodied. This is a sad state of affairs for which those truly interested in embodiment are partly to blame for often failing to specify the precise connection between their proposals and the body, and failing to describe what kind of bodies they have in mind. As a rule of thumb, any talk of bodily formatted representations belongs strictly to good old-fashioned computationalism, the corporeal adverb being superfluous. To point this out is only fair to researchers in the computational camp because their positions have never been naively unaware of bodily constraints when it came to concrete implementations, say, in classical robotics. The notion of embodiment must be revalorized. Intellectual honesty demands that any embodied theory should be able to provide precise answers two questions: What is its conception of bodies? What central role do bodies play in this theory different from the roles they play in traditional computationalism?

from scratch with the enactive approach (Varela et al. 1991; Thompson 2007), even though it saw one of its clearest formulations in Francisco Varela's later work (e.g., Varela 1997, 2000). The idea has roots in the earlier theory of autopoiesis (Maturana and Varela 1980), an attempt to give a systematic, generative, logical answer to the question: what is a living system? It also traces back to other notable precursors, as I will mention later.

If we take the project-of-the-world image at face value, then, albeit voided of any teleological implications, we get a hint of the kind of inquiry we are trying to circumscribe; ultimately one that offers important conceptual categories for any theorizing about cognitive phenomena. To ask about the meaning of this image, to ask how a medium projects itself into a subjects and objects, is to ask about the material conditions out of which pre-individual processes result in the individuation of living organisms, and the concomitant emergence of *their* world. It is also to demonstrate the intimate relation between these two moments, subject and world, as they co-emerge dialectically out of the same tensions found in pregnant materiality (see, e.g., Grosz 2011). It is also to ask in what ways these conditions relate to forms of psychic and collective individuation. Finally, it is to ask whether these material conditions provide only a background of enabling factors, which can then be assumed invariant across different instances of cognition, and therefore "safely ignored" for specific research projects, or whether, on the contrary, these conditions permeate all cognitive and social phenomena and make their understanding inescapable for any scientific project concerning the mind, no matter how specific.

LIFE AND MIND CONTINUITY

The enactive insistence on the continuity between life and mind has often been met either with impatience or misunderstanding. It is one of those situations where language can fail, giving the impression that one is talking nonsense or else saying something trivial and widely accepted. This means we must go back into it and attempt once again to clear the ground.

It is true: to say that there is continuity between life and mind could be seen as trivial and unimportant, especially in the context of widespread belief in the unity of science. After all, we do not know of any empirical instance of mental phenomena that does not also involve at least one living organism. The question is, does this knowledge matter for attempting to explain specific mental phenomena? Will attention to life-mind continuity have an influence in, say, theories of perception? Even to say that continuity implies that certain explanations used for understanding life will play important roles in any attempt to understand the mind could be met with shrugged shoulders. Do we not after all in disciplines like neuroscience, ethology, psychology, psychopharmacology, psychiatry, etc., already lean strongly on biological knowledge for support in explaining mental phenomena?

So one reaction is: we have been doing life-mind continuity all along. The other reaction is: this is nonsense. Defenders of the enactive approach may be at fault to some extent in this case. Synoptic formulas such as life = cognition have had their provocative initial impact diluted by their implausibility according to reasonable interpretations. Do such statements mean that life is coextensive with cognition? Are we performing a cognitive operation when, say, we digest our lunch? Are all currently living species equally "cognitive" since they are all equally alive? Is psychology reducible to biology? These and similar questions can tire the enactivist, but they are only fair if she is seen as standing behind a notion of continuity as the conflation of psychic and biological phenomena.

Between triviality and nonsense lies a deeper meaning of the continuity thesis. In the fewest possible words: mental phenomena constitutively demand explanations of individuality, agency, and subjectivity, and the principles and conceptual categories for these explanations are the same as those required by attempts to explain the phenomenon of life. Moreover, those conceptual categories and principles are not incidentally useful, but lie at the core of the question we have raised earlier: what are bodies?

Another way to put this is to say that a continuity thesis underlines the naturalistic project of the enactive approach. This is comparable to the same attitude adopted by John Dewey in his naturalistic theory of logic. According to him the primary postulate of such a theory is the "continuity of the lower (less complex) and the higher (more complex) activities and forms" (Dewey 1938, p. 30). Dewey maps the contour of the notion of continuity by making explicit what it excludes: a "complete rupture on one side and mere repetition of identities on the other; it precludes reduction of the 'higher' to the 'lower' just as it precludes complete breaks and gaps" (Dewey 1938, p. 30). Take the example of biological development; we cannot say in advance:

> that development proceeds by minute increments or by abrupt mutations; that it proceeds from the part to the whole by means of compounding of elements, or that it proceeds by differentiation of gross wholes into definite related parts. None of these possibilities are excluded as *hypotheses* to be tested by the results of investigation. What *is* excluded by the postulate of continuity is the appearance upon the scene of a totally new outside force as a cause of changes that occur. (Dewey 1938, p. 31)

To this we would add not so much an emphasis on "forces" outside the naturalistic framework but the rejection of the sudden appearance of fully independent novel levels of description—for instance, the realm of human normativity—without an account of how their emergence and relative autonomy is grounded on (understandable in terms of and interaction with) phenomena at other levels. This is as much a causal/historical point as it is ontological. The continuity thesis therefore proposes the need for a theoretical path that links living, mental, and social phenomena. The project, however, remains non-reductionistic for these three reasons: (1) it seeks explanations of emergent phenomena through theoretical and experimental investigations of, for example, self-organization and complex multi-scale interactions; (2) it replaces the notion of an independence of levels of inquiry (e.g., biology, psychology, sociology) with a notion of relative autonomy and postulates the conditions by which this autonomy can be tested;

and (3) it advances the possibility of various kinds of interactions between levels leading potentially to evolving forms of cross-level mutual dependence and transformations.

We have said it on other occasions: the enactive approach is best described as a non-reductive naturalism (Di Paolo et al. 2010) and the life-mind continuity is its core methodological, epistemological, and ontological attitude. Living and mental phenomena belong to intertwined branches of a same ontological tree (not that they must be equivalent or coextensive) and their study demands related epistemological tools.

DO THE SCIENCES OF THE MIND NEED AN ACCOUNT OF INDIVIDUATION AND AGENCY?

Theories of cognition should be able to provide the operational conceptual categories with which to describe their objects of study and distinguish them from those outside their remit. They should be able to say in concrete terms what sort of system, event, or phenomenon counts as cognitive and in which cases it does not. Accounts that do not meet this mark are pre-scientific. This does not mean they cannot lead to interesting or important knowledge or even to practical solutions to problems or technological innovations. It only means that the bits of knowledge so generated are provisionally held together by intuition or tradition and not by an articulated theoretical framework.

The conceptual categories mentioned earlier—individuality, agency, and subjectivity—lie at a blind spot of functionalist approaches to cognition, whether classical or "embodied." Such approaches must assume these notions as given and unproblematic. Otherwise, they cannot work. Let us examine why.

The idea that it is possible to explain cognitive phenomena in terms of the commerce of functional, representational neural states, bits of information, vehicles and content, etc. implies that a certain stationarity[3] is needed in the permitted variations of states that the cognitive system may undergo. We can call this the informational frame within which functional states have well-defined roles. By definition of what it means to be an information-processing system, the cognitive machinery that processes information cannot therefore change in non-stationary and open-ended ways without at the same time limiting the range of applicability of a functional explanation.[4] One solution adopted for dealing with this problem, say, in theories of learning, is to assume that cognitive systems operate in at least two sufficiently distinct time scales: a fast time scale that corresponds to a settled functional system, and a slow time scale that corresponds to

---

[3] A process is said to be stationary if the probability distributions for its states do not change over time, that is, if it does not present transient trends that alter general statistical properties such as mean or standard deviation.

[4] It may still be possible for non-stationary changes to occur within a complex system such that certain regions of the system conserve a relative stationarity. Within these regions, at the appropriate time scales, it may be possible to perform valid functional analyses at a local level.

how this system changes its functional frame over time. The interaction between these time scales is hierarchical and essentially non-messy. They accord with Simon's (1962) postulates for near-decomposable systems. The whole framework remains functionalist in that changes in the frame occur in a way that is itself given by a more encompassing stationary frame, for example, plasticity rules in a neuronal network.

But here we face a problem: the question of what constitutes a cognitive system as an extended spatiotemporal entity (in essence, the issue of how it *becomes* a cognitive system) as well as the question of how a domain of significance is constituted in the here and now of a concrete situation (e.g., how activity is framed as appropriate to a context or a motivation)—two intimately connected questions in the enactive approach—demand answers in terms of transformative (frame-changing, frame-establishing) processes, i.e., they demand a non-stationary story. Given these constraints, only two options are open for functionalists: either (1) assume that the issue of becoming a cognitive system is a non-question, i.e., that nothing except convention distinguishes cognitive systems from any other system of functional relations, such as toasters, or (2) assume that it is interesting but irrelevant, i.e., that the answer to this question corresponds to a different science, such as biology, and that once given, it does not contaminate cognitive science and one can safely assume that it does not bear on the explanation of concrete cognition. Similar options are available for the second question, that of the emergence of concrete frames of significance.

Functionalism, even in its embodied versions, has (mostly tacitly) gone for either one of these options. The functionalist would be safe if there were nothing special about cognitive systems that would distinguish them from other systems that could be assumed to be stationary. Alternatively, even if during a period of construction the cognitive system does not verify the assumption of stationarity, this could be assumed to be a well-delimited period of transient transformations outside the remit of cognitive science, after which, the cognitive system can be safely be treated as stationary.

Why is stationarity at odds with an account of individuation? There is, first, an empirical answer, namely that such seems to be the nature of all known forms of cognitive systems: they grow, develop, adapt to unforeseen circumstances, and seem to have an open-ended (though not unconstrained) reserve of potentialities, which we have no reason to assume are all pre-given at birth, since potentialities are always relational with respect to an open-ended environment. As the enactive story unfolds, a stronger, conceptual answer emerges. It postulates that ongoing, open-ended, precarious processes are logically necessary for what makes a system cognitive. Like living systems, cognitive systems are identifiable as centers of activity and perspective. Cognition occurs when there is a cognizer that cognizes about something. This means that there is an entity that takes a stance, and from this stance relations between itself (the cognizer) and its world are inherently meaningful. But there cannot be any such relations unless the entity we call the cognizer is also an individuated entity. And as we will see, these relations cannot be meaningful unless individuation is an ongoing, open, precarious process; i.e., a non-stationary one. The possibility of unpredictable, frame-transforming changes is inherent to being a cognitive system, even in the particular circumstances where these changes are not actually occurring. Hence, to be a cognitive entity is to be a (generally) non-stationary organization in a (generally) non-stationary relation with the world. Since functionalism is limited to cases in which we can safely make the stationary approximation, it follows that it cannot account for fundamental aspects of cognition.

This is not merely an arcane conceptual issue. In many ways, its implications are always close to the surface in concrete research. When we study attention, volition, sense of agency, decision-making, value systems, learning, etc., all of these aspects are implicit. What makes a cognitive system one that can act purposefully, do so with spontaneity, have concerns about its ongoing well-being and activities as well as concern for others, decide correctly, recognize and solve problems, and so on? To try to answer how such acts are performed without understanding why they are carried out at all, why they are of any relevance for the cognitive agent in the first place, does not even amount to half the story. In the mind sciences, there can be no general account of "how it works" without also offering an explanation of "what is at stake and for whom," since these questions are inseparable. Otherwise, we are speaking of complex systems theory, not cognitive science.

Without a solid account of individuality, agency, and subjectivity, we have not even scratched the surface of a theory of the mind, and all the well-established results are provisional because we have no theory that specifies their range of validity, only intuitions and empirical data (which can only give instances of (non-)contradiction of an assumption, but not in themselves explicate the limits of its generalization).

Similar points can be made about the inexistence of a theory of agency in cognitive science, both traditional and "embodied." Again, nothing in functionalism, except external convention or convenience, enables us to theoretically distinguish between a system that is simply coupled to its environment, like the planets in the solar system are gravitationally coupled to the sun and each other, and a cognitive system that is an agent in a meaningful world. An agent does things as well as has things happen to it. Again, in practice, this lack is always complemented by some tacit commonsense assumptions when focusing on specific research concerns. We tend to assume that there is a clear difference between a person moving an arm of her own volition vs. having it moved by an experimenter. But do we have in principle ways to distinguish less obvious cases or to question whether accompanying the experimenter's movements and not opposing them is not also a volitional act?

In short, it seems that there are good reasons to bring to the surface some of the hidden assumptions of the prevalent functionalist framework in the sciences of mind—not only as a healthy exercise, but in order to offer a possible explanation of why certain questions have never been the center of cognitive science research, such as the question of cognitive becoming or the question of the constitution of agency. Enactive theory has, in addition, deeper reasons. These are the issues that permeate all aspects of cognition for this approach. However, this does not mean that it is not possible sometimes to assume that some of these theoretical worries will have limited impact in specific cases of interest. Whether this is a good epistemological move or not, however, necessitates a theoretically loaded framework to justify it, very much in the same way as the theory of relativity itself provides the justification of what conditions validate the applicability of Newtonian mechanics.

We then turn to reviewing these deeper reasons in the next sections.

AUTOPOIESIS

The enactive view of life and mind derives from the theory of autopoiesis—if by derivation we mean the historical sense of progression of ideas and not the logical sense of entailment. In fact, much of what is predicated by enactivists, especially in relation to norms, agency, and social interaction, is different and even quite at odds with classical autopoietic theory (Maturana and Varela 1980; Maturana 2002). I will not rehearse the technical arguments but will highlight some of these differences as we proceed.

The theory of autopoiesis emerged in the 1970s as a response to prevailing views in biology, neuroscience, and psychology, which lacked deep scientific conceptions of organisms, agents, or persons. Autopoietic theory rejected notions of information processing, since they tend to conflate phenomena at the supra-, sub-, and organismic levels. Instead, it adopted an epistemology based on systems theory to postulate that the identifying feature of any system is not its conventional labels, nor its contingent spatiotemporal arrangements, but its organization. Define the organization of a system and you will achieve two things. First, you will define a class of systems that share this organization, such that different instances in this class can be said to belong together, regardless of how much they differ in terms of how their organization is actually instantiated. This seems obvious enough. However, any predicate based on a logical analysis of the properties of a given organization, will ipso facto apply to all instances of a class, which is a useful conceptual tool when we deal with complex systems. Second, given a particular instance of a class of systems, the conservation of its organization is what permits an observer to postulate its identity through time as that particular instance; its haecceity. Thus, while the observed system may change in structure, an observer can say that this is the same system provided its organization is unchanged. Conversely, a change of organization is sufficient to say that the same spatiotemporal arrangement of processes is no longer the same system that it was before. It has transformed into something else or it has disappeared.[5]

It is possible to raise some criticisms at this stage already, for instance, the apparent lack of material and temporal constraints underpinning the idea of a sustained

---

[5] Classical autopoietic theory works like mathematical set theory: while a given organization may be instantiated in different concrete structures, it is also the case that a particular concrete system may embody more than one organization, on condition that there is no contradiction between them. Thus *this* dog is a mammal, but also an animal and also a living system, each category implying a broader organization and class identity.

organization. Notice also the silence about history and about relations; strictly speaking, all that needs to be known about the identity of a system is given intrinsically and contemporaneously from within the system itself. It is not tendencies or relations or forms of becoming or potentialities that will take ontological precedence in the theory of autopoiesis, but conserved being as defined by organizational properties. The latter may themselves be relational but they belong to the system in that they are the sufficient constitutive relations that establish the class identity of the system. They are to be distinguished from relations *between* systems so defined.

Many conceptual categories that are historical or relational cannot be approached in this way: e.g., being the offspring of, being a twin, being capable of reproduction, belonging to a group, etc. Such relational, historical observations, according to autopoietic theory, belong to the cognitive domain of the observer, which is a quick way of dismissing them in order to focus only on set-theoretic, immanent, operational relations as they are actualized here and now. These, however, are no less observer-dependent since they always imply an epistemic grounding on the part of the observer. In practical terms, the choice of what counts as relevant variables, parameters, and constraints is always precisely that: a choice. Conceived at the level of quantum interactions or at a scale of attoseconds, a living cell does not reveal itself to an observer as autopoietic. Exactly the same epistemic condition holds for the relational/historical categories that autopoietic theorists feel uncomfortable with—once the domain of observation is chosen, their manifestation in this domain is not arbitrary nor is it a matter of convenience or convention.

These worries will resurface later. In any case, based on this systemic framework, to ask the question of how an organism, a living system, should be conceived of is to ask about its organization. More precisely, the central move in autopoietic theory is to propose the description of the organization of a class of systems such that living systems fall neatly into this class, and nonliving systems fall outside it. The proposed description is the definition of autopoiesis as a network of molecular processes undergoing material transformations (including production and destruction) with the following organizational conditions: (1) the network realizes the relations that give rise to the production of its constitutive processes, and (2) the processes constitute the network as a concrete unity in space and define its topological relations (see Maturana and Varela 1980, pp. 78–9; Varela 1979, p. 13).

We may call Condition 1 the *self-production* condition and Condition 2 the *self-distinction* condition.

“Autopoiesis is necessary and sufficient to characterize the organization of living systems” (Varela 1979, p. 17; Maturana and Varela 1980, p. 82, emphasis removed). It is a set-theoretic description of a self-producing and self-distinguishing system, an idea with precursors such as Kant’s conception of life as the mutual production of parts and whole and Hegel’s reworking of this idea (Kreines 2009; Michelini 2012), the work of physiologists in the Revolutionary period like Xavier Bichat (1800), and mid-twentieth-century philosophers like Hans Jonas (1966) and Georges Canguilhem (1965) (who already in 1951 used the term *autopoietic* to refer to the character of organic activity).[6] Autopoiesis is the idea that a living system is one that is constantly constructing itself and by this activity making itself distinct from its environment.

Introducing an operational description of the organization of living system is only the first step because, as we said earlier, anything that can be deduced logically from this definition will apply to any system that shares this organization. In this way, autopoietic theory challenges various widely held postulates, such as the very possibility of mental representations, and offers non-traditional conceptions of communication, sociality, evolution, and language (Maturana and Varela 1980; Maturana 2002). It would take too long to review these implications here.

The force of some of these challenges to tradition is inherited by the enactive approach, which nevertheless questions some of the starting assumptions and interpretations of autopoietic theory.

The Co-Definition Between Organism and Environment

Worries about autopoietic theory cannot hide the fact that the enactive approach is a historical offshoot of these ideas, in particular, as exemplified in the later work of Francisco Varela. Already moving beyond some of the limitations mentioned earlier, Varela conceived of autonomy as an idea wider than autopoiesis, applicable to other phenomena exhibiting some form of self-sustained identity through operational closure. In other words, the conditions of self-production and self-distinction could be applied in several different domains, not just the domain of molecular transformations (Varela 1979). Eventually, his contributions to work on immune networks cell assemblies in neuroscience, and his contact with various traditions exploring human experience (phenomenology, meditation, psychoanalysis) led him along a path of refinement and reformulation of some of the assumptions of autopoietic theory (e.g., Varela 1997, 2000; Varela et al. 1991).

There are some important differences between the enactive and the strictly autopoietic conceptions of life. These differences involve dropping some conceptual barriers such as the one separating constructive (internal) and relational (cognitive) phenomena. They also involve the introduction of precariousness as a constitutive element eliminating trivial interpretations of autonomy and necessitating a notion of adaptivity (Di Paolo 2005; Di Paolo and Thompson 2014). Perhaps the most crucial difference is an idea proposed by Varela explicitly toward the end of his life (Weber and Varela 2002): that autonomy as a

---

[6] "Only after a long series of obstacles surmounted and errors acknowledged did man come to suspect and recognize the autopoetic [sic] character of organic activity" (1965, p. 9). The text on experimentation in animal biology was part of a presentation given by Canguilhem in 1951 in Sèvres, France, and published as part of the collection *La Connaissance de la vie* in 1952 and re-edited in 1965.

property of living and cognitive systems is the grounding of the first layer of teleology and normativity according to which an organism can relate to the world in meaningful terms. In other words, adaptive autonomy, a non-mysterious property of some natural systems, is the condition of possibility, as well as the conceptual ground of sense-making, or simply, mind.

Instead of rehearsing this story, which has been told on other occasions (Di Paolo 2005, 2009; Thompson 2007), I would like here to present the enactive concept of life in a different form. I will approach it as the overcoming of dialectical tensions in the relation between organism and environment.

This relation is indeed where all the fascinating stuff happens, at least for those interested in cognition. The state-determined dynamics of the internal processes that constitute an autopoietic system have been used in the original literature to argue against the possibility of mental representations. In a nutshell, whatever impinges on an autopoietic system has only a triggering, not a formative effect on what subsequently goes on internally within the system. This is determined solely by the internal dynamics (Maturana and Varela 1980, pp. 81, 127). For this reason, all that is possible to say about the internal processes and the relations that the system enters into as a whole is that they have a relation of coherence, meaning that these conditions are not at odds which each other; otherwise the system would stop being autopoietic.[7] This is what is called structural

---

[7] Because it does not bear on the main thrust of this section, I mention only as an aside an overlooked problem with this kind of reasoning. The notion that the determining role in a state-determined system is played only by the states of that system and that an external perturbation at most triggers a particular chain of internal changes seems quite straightforward from an abstract perspective. And yet we should ask whether this is sufficient for claiming that the triggering perturbation is incapable of playing a formative/structuring role in the system's subsequent changes of state. It obviously could not create these changes all by itself, and the autopoietic argument against a *determining* effect of an external stimulus resulting in a representational token "shaped" or informed solely by the properties of this stimulus still stands (all that matters for this is that internal dynamics are state-*dependent*, not state-*determined*). However, once we consider complex systems far from equilibrium that are materially and energetically open to exchanges with the environment, we must also take into account that such systems can reach states with a variety of types of stability, including critical metastable states which are poised between a few lower energy, relatively more stable options. For some types of complex systems, the evolution toward such critical states is actually likely (Bak et al. 1987). These critical states by themselves, save for fluctuations, will not evolve into any of the lower energy options unless presented with an external trigger. This is precisely the meaning of information used by Gilbert Simondon (2005) in his analysis of individuation processes in nature. One of his recurrent examples is the process of crystallization. A supersaturated solution (the system) will remain in a metastable state until the process of crystallization is initiated by the presence of a seed (the perturbation). During this process there is a passage to a lower energy state "shaped" or in-*formed* by physical characteristics of the seed. Some compounds can potentially crystallize into more than one lattice shape—which lattice is actualized will depend on the seed. In this sense, a trigger is indeed a perturbation and does not in itself carry out the process of crystallization or the determination of the possible options for crystal structures. These virtual alternatives are defined by the system's critical state and the energy and material resources for formation of crystals are also provided by the system. Depending on the particular question of interest, however, it may make sense to speak of an *in-formational* process (Oyama 2000) that occurs at the *encounter* of seed and solution involving not abstract form being bestowed on formless matter (as content on vehicles), but material form (the seed) and being propagated (transduced) in interaction with both actually and virtually pre-formed matter (the critical-state solution).

coupling. The interpretation of the relation between internal and external phenomena that emerges from taking structural coupling as fully characterizing the organism-environment relation is a strong one:

> Systems as composite entities have a dual existence, namely, they exist as singularities that operate as simple unities in the domain in which they arise as totalities, and at the same time they exist as composite entities in the domain of the operation of their components. The relation between these two domains is not causal; these two domains do not intersect, nor do the phenomena which pertain to one occur in the other. (Maturana 2002, p. 12)

There is no possibility of naturalizing cognitive phenomena by using only the notion of structural coupling. It specifies only a minimal condition on the organism–environment relation: an organism is alive as long as it does not enter into destructive relations with the environment. But this minimal condition is insufficient to understand the inner links between life and meaning. In line with the principles of continuity discussed at the beginning, we must seek more specific organism–environment relations that allow—without contradicting structural coupling—a conclusion different from the dualism of non-intersecting domains.

Here it is worth quoting Varela at length in a passage written in 1996 for the preface to the second Spanish edition of the canonical text *De Máquinas y Seres Vivos*:

> One of the criticisms that could be made of this work (as well as of my 1979 book), is that the critique of representation as a guide for explaining cognitive phenomena is replaced by a weak alternative: externality as a mere perturbation of the activity generated by operational closure, and which the organism interprets, be it at the cellular, immune, or neural level. To replace the notion of *input-output* with that of structural coupling was an important step in the right direction because in this way we avoid the classical language trap of making the organism into an information-processing system. But it is a weak formulation because it does not propose a constructive alternative since it leaves interaction in the fog of being a mere perturbation. Often, the critical point has been made, that *autopoiesis*, as formulated in this book, leads to a solipsistic position. Because of what I have just said, I think this criticism has some merit. The temptation of a solipsistic reading of these ideas derives from this: the notion of perturbation during structural coupling does not adequately take into account the regularities that emerge from a *history* of interaction in which the cognitive domain is neither constituted internally (in a way that effectively leads to solipsism) nor externally (as in traditional representational thought). In these last few years I have developed an explicit alternative that avoids these two stumbling blocks, making of historical reciprocity the key of a *co-definition* between an "autonomous" system and its environment. It is what I propose to call the perspective of *enaction* in biology and cognitive science. (Varela 2000, pp. 447–8, original emphasis, author's translation)

It seems that the canonical picture of non-intersecting domains, one corresponding to the phenomena participating in self-construction, the other corresponding to relations entered into by the organism as a whole, is misleading. It is true that the lack of a determining relation between phenomena in one domain on phenomena on the other means that these domains are irreducible to each other. Thus, if we are in possession of all the relational facts between organism and environment, this knowledge will not suffice to predict with certainty how the internal dynamics of the organism will unfold. Conversely, knowledge about all the neural and physiological facts is insufficient for ascertaining the behavioral and perceptual facts likely to ensue in the relational domain. In each case there is a remainder of determination provided by each domain on itself. This is an important systemic lesson, but all it says is that one cannot reduce one domain to the other, not that the domains are non-intersecting, not that they cannot enable and constrain each other. A history of influences between phenomena in each domain may indeed lead to more intimate (partial) co-determinations than simply the minimal mutual tolerance of structural coupling.

Let us consider again the two conditions of the definition of autopoiesis (self-production and self-distinction), paying special attention to what they imply with respect to the organism–environment relation.

The self-production condition specifies that the network of component processes realizes the relations that give rise to the production (or regeneration) of the same processes. Realizing such relations in the real world implies establishing the conditions by which the flows of matter and energy present in the environment can be used in the regeneration of metabolic processes. Let us for a moment imagine what would be the ideal situation in which a form of life could realize these relations. This situation would correspond to the idealized circumstances in which every possible encounter of the organism with the external world produced a positive contribution to autopoiesis and none produced a negative contribution. In other words, if we take self-production on its own, the ideal condition would be one of total openness, such that every possible flow of matter and energy is taken advantage of. Unrealistic as this is, the case is that no relation with the environment would facilitate self-production more than this one if it were possible.

Consider now the self-distinction condition: the autopoietic system constitutes itself as a well-delimited unity with specific topological relations. What would be the relation with the environment that would most ideally realize self-distinction? One of total robustness to any environmental influence, i.e., perfectly shielded boundaries protecting the system. In this case, no possible interaction with the world could possibly put at risk the condition of being a distinct unity, simply because no interaction with the world would have any effect on the system.

There is a primordial tension to this definition of life insofar as the organism-environment relations that best satisfy each of its two conditions tend in exact opposite directions. The tension is well captured by the original split of the autopoiesis definition in two separate conditions. The organism must tend to be self-enclosed to assert its distinctiveness as an individual, but it must also tend to be open to sustain its self-production as a far-from-equilibrium system. In the classical literature, this tension is not further thematized. It is apparently resolved by an offhand clarification: operational closure does not imply material or energetic closure. But where in the world can we expect matter and energy to always flow in or out of a system in the "abstract," i.e., with zero influence on organizational/structural relations? This separation of function and flows is an abstraction aimed at in technological applications. Fuel is supposed to provide pure energy for the car engine and not alter its function. But we know that even in a system specifically designed to approximate this condition as much as possible, this is only an idealization (witness the effects of low octane fuels on uncontrolled ignition in the combustion chamber, leading to engine "knock" and eventually to serious engine damage). In biology especially, most of the matter that flows across the cellular membrane is pre-formed (high-energy compounds, proteins, plasmids, etc.)

It is problematic to say that matter and energy may flow freely across the boundary of the organism, because if this happened it would soon become a violation of the self-distinction requirement due to uncontrolled transformative effects. On the contrary, pre-formed active matter and energy can only flow *conditionally* across the organismic boundary.

Let us clarify how. The ideal organism–environment relation for each requirement in the definition of autopoiesis negates the other requirement. Given this pull of opposites for the organism–environment relation, there is one solution, which is the dialectical overcoming of this tension. A real-world autopoietic system would also need to be a dynamically adaptive one, which by necessity would be open to selected environmental flows (e.g., those that contribute to the condition of self-production) and closed to others (e.g., those that act against the condition of self-distinction). These options are presented schematically in Figure 4.1.

The overcoming of the primordial tension of autopoiesis takes us closer to the enactive conception of life. As we see, this conception is derived from autopoiesis, but provides an alternative interpretation of the possible relations between the constructive domain of production and regeneration, and the domain of organism–environment coupling. These irreducible domains are no longer interpreted as non-intersecting. On the contrary, assuming precarious, far-from-equilibrium conditions, the two requirements of autopoiesis demand opposing ideal situations; they "pull" in different directions as tendency and counter-tendency. This tension is managed over time and the internal and interactional domains relate in ways defined by the adaptive capabilities of the organism. Matter and energy flows do not provide abstract and form-less raw resources; they contribute to sustaining but also possibly modifying the way autopoiesis is realized. These effects are more salient as we observe a history of interactions.[8]

---

[8] Some of the transformative effects of environmental intercourse on processes of self-individuation are studied in origins of life and early evolution modeling. For instance, Froese, Virgo, and Ikegami (2012) explore possible routes to the origins of metabolism in a model of autocatalytic individuation in an excitable spatial medium. The authors show how reaction-diffusion self-individuated patterns become mobile, even in the absence of chemical gradients, when the spatially individuated pattern corresponding to one autocatalytic reaction "incorporates" another pattern formed by a different autocatalytic reaction. The resulting system becomes spatially asymmetric and spontaneously mobile. It is also well known that bacteria can exchange plasmids (small DNA molecules) horizontally, and even the consumption of certain chemicals can alter metabolic pathways and overall functionality,

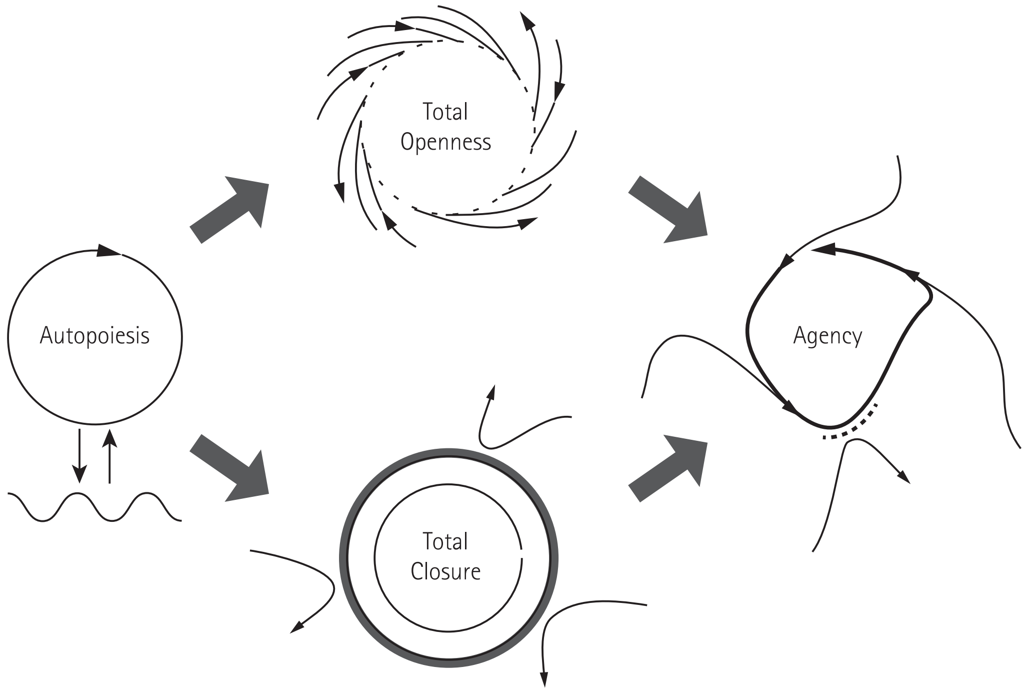

FIGURE 4.1. The primordial tension of life. An autopoietic system is represented on the left by a circle with an arrow closing on itself. The environment is represented as a wavy line and flows of pre-formed matter and energy in coupling with the autopoietic system are indicated with arrows. Under precarious conditions, each of the two requirements for autopoiesis (self-production and self-distinction) leads separately to opposing ideal relations with the environment. If all possible environmental flows could be used for self-production, this would entail an indistinct open system (top). In contrast, a system shielded from environmental flows would be optimal for self-distinction, thus resulting in full isolation (bottom). The organism would not be alive in either of these extreme cases (this is depicted by the open circles). The ideal organism–environment relation for each requirement negates the other requirement, which is why in neither extreme condition can both requirements be met simultaneously. On the right we see the dialectical overcoming of this primordial tension: an adaptive autopoietic system—an agent—able to distinguish and regulate flows that contribute to self-production and self-distinction and avoid flows that act against these conditions. The primordial tension is actively regulated over the time domain. Agency is therefore entailed in any material realization of autopoiesis.

Copyright © 2017 Ezequiel A. Di Paolo, Thomas Buhrmann, and Xabier E. Barandiaran, with permission. Originally published in *Sensorimotor Life: An Enactive Proposal*, Ezequiel Di Paolo, Thomas Buhrmann, and Xabier E. Barandiaran, Figure 5.6, p. 135, Oxford University Press, Oxford, 2017. https://global.oup.com/academic/product/sensorimotor-life-9780198786849). Reprinted here with permission of the authors.

The enactive conception of life expresses in systemic terms the characterization of metabolism offered by Hans Jonas (1966). In looking at the relation of life to matter, he describes it as a "dialectical relation of *needful freedom*" (Jonas 1966, p. 80). He does not specify the need for an adaptive regulation through which the organism evaluates its

for instance, in energy-taxis in bacteria. Egbert, Barandiaran, and Di Paolo (2012) coined the term "behavioral metabolution" to describe some of the evolutionary implications of these processes of transformation in metabolism and behavior through the incorporation of pre-formed matter.

coupling to specific environmental flows, but the need for active regulation is implied in his recognition of the primordial tension of life: materials are essential to the living organism but its identity is dynamic, not tied to the individuation of material constituents but emerging instead as the (risky) ongoing adventure of "riding" material changes "like a crest of a wave" and "as its own [the organism's] feat" (Jonas 1966). This helps us highlight a difference between classical autopoiesis and the enactive view of life. Unlike the former, life in the latter is inherently dynamic and inherently "at risk" because the overcoming of the primordial tension is an ongoing achievement (see also Froese and Stewart 2010, for whom the more static conception of life is connected to the influence of cybernetics on autopoietic theory). Life in the enactive approach is always in a transient, not just empirically, but constitutively as the only way of managing the primordial tension between its own opposing trends.

We should notice that the intrinsic dynamical and relational nature of the enactive conception of life as adaptive, precarious autopoiesis is problematic for functionalism. It is thanks to the ongoing and risky engagement and regulation of the coupling with the world that any identity is sustained and any interiority possible. There is no stopping at stable states in this view of life, no equilibrium, and due to the unpredictable nature of environmental relations, no guaranteed stationarity either. Functional analysis applied to life and mind can only be valid as a limiting case, not as general theory.[9]

Projecting a World

If we add to the minimal condition of structural coupling a condition of adaptive regulation of the organism–environment interaction, as seems necessary for the overcoming of the primordial tension of autopoiesis, then we are on the road toward articulating the historical co-definition between organism and environment that Varela speaks of.

First, it is important to notice that these moves do not imply abandoning the naturalistic approach, nor do they imply sneaking in teleology within a systemic description (as Villalobos 2013 worries). Quite the contrary, by introducing the notion of adaptivity, the enactive approach proposes a bona fide naturalization of teleology and normativity (Di Paolo 2005; Thompson 2007) as well as the notion of agency (Barandiaran et al. 2009). Naturalization would not be valid if we assumed teleology and normativity in the starting postulates, but would fail if these did not appear at some point in our story.

Second, it is also important to highlight that by co-definition we are not simply talking about historical correlation; i.e., the mere concatenation of state-dependent changes whereby the current state depends on the past history of internal changes that have occurred in the autopoietic system, which of course are historically related to

---

[9] For an enactive/organizational approach to biological function, see the work of Alvaro Moreno and colleagues (e.g., Moreno 2010; Nunes-Neto et al. 2014; Saborido et al. 2011). We should also mention that regulation could act with the effect that some subsystems are able to adaptively sustain a given function or repair it when partially lost.

perturbation events and their internal compensation (e.g., Maturana and Varela 1980, p. 102). The notion of co-definition entails stricter conditions.

To see these points more clearly we need to consider adaptivity in more detail. Adaptivity (Di Paolo 2005, p. 438) is operationally defined as a system's capacity, in some circumstances, to regulate[10] its states and its relation to the environment with the result that, if the states are sufficiently close to the boundary of viability:[11]

1. Tendencies are distinguished and acted upon depending on whether the states will approach or recede from the boundary and, as a consequence,
2. Tendencies of the first kind are moved closer to or transformed into tendencies of the second, and so future states are prevented from reaching the boundary with an outward velocity.

Here we should notice that adaptivity relies on a broader dynamical systems ontology than the one used by autopoietic theory. The latter leans strongly on the notion of a state-determined system according to which the state of any variable in the system is determined only by the previous state of this and other variables in the same system, not by anything external to it. Any external interaction enters into the system as a perturbation, to which the system responds by an internally determined compensatory change. This notion is compatible with the enactive approach but its application differs in two crucial ways. Firstly, while state-*dependence* is always valid, state-*determination* as such is only valid in mathematically autonomous systems, i.e., systems that among other things are not subject to time-dependent couplings. This is not necessarily always the condition in which a living system finds itself. Coupling with other systems introduces all kinds of time-dependent and quasi-regular forms of perturbation (e.g., the rhythms of daylight, the rhythms of the seasons) as well as the emergence in interaction with other organisms of stable patterns of relational dynamics underdetermined by

---

[10] Here, before the adaptivity requirements are mentioned, it would have been more appropriate to use the term "modulate" instead of "regulate" in the original definition. The difference is subtle. To modulate a system is to alter the conditions of its operation, for instance, by altering parameters or relations to other systems. To regulate a system is to modulate it according to some norm. It is correct to use regulation when speaking of adaptive systems and modulation when speaking in more general terms.

[11] There are different ways to measure viability according to the organism or model in question. In a model of mobile self-producing protocells, Barandiaran and Egbert (2014) distinguish not only viable and nonviable states in the space of essential variables that affect the system, but they also define precarious regions in this space as those states in which the system is not yet dead but in the absence of any environmental changes (e.g., external intervention) the trajectory of its states will inevitably cross the viability boundary. In a related model studying plastic transformation in self-individuated protocells, Agmon, Gates, and Beer (2015) measure the viability of a given configuration as the average number of perturbations it would take for the protocell to disintegrate. In this model, moreover, the authors demonstrate that adaptive transitions can occur in the absence of strict monitoring and regulation of changes in viability, but as an emergent aspect of transformations that tend to increase viability. These two models show two possible examples of how viability space can be modeled and measured.

the participants themselves, which constrain their individual operation (De Jaegher and Di Paolo 2007). The second difference is that enactive theory uses notions of dynamical landscapes and different kinds of stability to operationalize concepts such as tendencies and dispositions in dynamical terms (see Di Paolo 2015). In short, there is an acceptance of the concrete reality of potentialities, virtualities, dispositions, etc., which are articulated in operational concepts such as critical metastable states, gradients, flows, and dynamical landscapes.

A system able to adaptively regulate itself and its coupling with the environment is by necessity a system in which parameters and conditions of operation change over time. This being so, the notion of co-definition is not one whereby systems' *states* become historically correlated, but more strongly, one in which systems' *operating conditions* do. We are then likely to find between organism and environment a relation of mutual shaping, not just a relation of correlated states. In other words, adaptivity operates within the constraints of structural coupling (the organism conserves its autopoietic organization) but it can introduce through a history of interaction additional conditions and coherences between the transformation of both states and conditions of operation in organism and environment. These coherences are not just in the cognitive realm of the observer as autopoietic theory says, but operate as constraints and facilitative conditions that shape the structure of the organism and of the environment (including other organisms) through time. This is how the enactive approach can articulate operationally the notion of *co-definition* described by Varela in the passage quoted earlier.

Thus, adaptive interactions contribute to shaping the organism in fundamental ways. Durable structural/functional effects on the organism related to habit, history of use, and training are too numerous to mention. To name just one striking example, consider the case of underwater vision in Southeast Asian children. Moken children living in the Burma archipelago along the West coast of Thailand routinely dive in search of shells and clams without the use of visual aids. Their underwater acuity is roughly twice as good as that of European children. On land both groups do not differ in acuity or accommodative power. They differ, however, in pupil size: a Moken child's being significantly smaller, which helps to improve acuity underwater (Gislén et al. 2003). The difference is demonstrably associated with the regular performance of underwater visual activities: pupil size reduction and improved underwater acuity can also occur in Swedish children after only one month of training in recognizing visual patterns underwater (Gislén et al. 2006).

Many other clear examples of adaptive historical co-structuring involve social and collective systems. Consider horizontal gene transfer in bacteria whereby plasmids are exchanged that may be incorporated into chromosomes and alter bacterial resistance to antibiotics (Koonin et al. 2001). Or consider the epigenetic effects of mother rats licking and grooming their young offspring. If a rat mother fails to engage in grooming and licking of her pups during the first couple of weeks after birth, those pups will grow to have deficient regulation of acute stress responses (Liu et al. 1997). Licking and other forms of tactile stimulation promote the formation of glucocorticoid receptors (GR) in the pups. This increased expression of GR associated with different brain regions, including the hippocampus, closes a negative feedback loop that helps in the regulation of short-lived stress responses. Insufficient GR make this feedback loop fail and lead to sustained states of high anxiety (which in adult mothers are likely to promote poor contact with pups in the next generation).

The question is whether it is possible to provide explanations of these effects using the notion of structural coupling. The answer is no. The condition of non-lethality of interactions is necessary but insufficient to explain the regularity of these effects. We could simply imagine as viable, for instance, that pupil size in Moken children remains unaffected by frequent underwater diving. Instead, there is in these examples an undeniably adaptive shaping at play, with cumulative historical effects that lead to differentiation in the realization of self-production and self-distinction. An enacted relation with the environment leaves both external and internal historical traces that coherently alter both the domain of organismic construction and the domain of external interactions because each domain constrains and enables the other.

The intimate relation that develops historically between agent and environment is one therefore of *mutual shaping*, which is also manifested on the environmental side. Autopoietic theory puts most of the emphasis on one arc of the organism–environment coupling, the perturbations that impinge on the autopoietic system. But the enactive approach—one could say this is one of the defining features, the one that gives the approach its name—emphasizes the role of action and involvement in the world. The emphasis is on the whole organism–environment coupling (sensorimotor or otherwise).

It is in the environmental consequences of living activity that the organism objectifies its sense-making, both for itself and for other organisms. We see this clearly in the wide variety of examples of niche construction in biology (Odling-Smee et al. 2003), in particular as they involve spatial constraints that enhance historical relations through the path-dependence brought forth by locality (Silver and Di Paolo 2006). Both the increasingly self-differentiated subjective world of the sense-maker and the increasingly objectified properties of the environment partially lose their mutual externality as organisms act (and eventually labor) in a transforming relation of productive activity (see also McGann 2014 for discussion about the need to elaborate an enactive theory of collectively shaped environments).

A history of mutual structuring can also lead to novel forms of extended autonomy. Such is the case of extended physiological circuits in hermatypic coral, the trapping of air bubbles that allow insects and spiders to swim underwater, or the construction of sound amplifying burrows by mole crickets (to name a few of the fascinating examples discussed in Turner 2000).

The picture of mutual co-definition between organisms and environment is even more compelling when we consider life as originating in communities from the very beginning, an issue that we have not discussed here and would deserve a more thorough separate treatment.[12] If environments can be the source of structuring powers, which the organism can to some degree adaptively select to be open or closed to, this is a fortiori the case if we take account of the collective nature of life. Here not only do we find organisms interacting with structuring/structured flows of active matter and energy available in the inorganic world but with objectified biological and historical products, sedimented practices and acts that play the role of signals, symbiotic relations, and even whole other organisms. Historicity is fueled by numbers, as there are clear material and temporal limitations to the effects that a single organism may have in transforming its world.

Conclusion

Returning to Merleau-Ponty's formulation, which we deliberately pushed in a direction that was latent in its language, we can say that a subject projects a world and is itself projected by it, in virtue of how sense-making is constituted by adaptive, historically shaped, organism–environment relations. It is on this world that subjects depend for their continuous existence both as experiencing beings but also, more fundamentally, as living bodies. Varela's long quote echoes this view by speaking of a relation of co-definition.

Our examination of the enactive conception of life, however, emphasizes two aspects that are not explicit either in Varela's or in Merleau-Ponty's formulations. The first is that the constitutive precarious conditions of all life, without which it would not exist as such, demand an ongoing process of organismic individuation which is primordially at odds with itself and as a consequence can only surpass its own tensions dialectically by adaptively regulating its relations with the environment from which it emerges. Sense-making is precisely the opening into the temporal/historical dimension in which viability is made possible by time-managing otherwise unsolvable contradictions. At the fundamental level, enactive bodies are constantly buying time. The adaptive relation, the basis of all forms of cognition—insofar as cognition implies time-oriented subjects

---

[12] The question of whether some form of fundamental collectivity is implied in the enactive conception of life is an important one, but remains so far unresolved. It seems, at first, plausible that this is not the case. However, this view is based only on the apparent conceivability of the emergence of a singular organism without entering into the detailed conditions of feasibility of such an emergence. It may well be the case that it is impossible to conceive of life arising singularly once these conditions are taken into account. For instance, studies in synthetic biology are beginning to pay attention to the effects of collective protocell interactions on the formation of prebiotic lipid vesicles (e.g., Shirt-Ediss, Ruiz-Mirazo, Mavelli, and Solé 2014). The very conditions for life to exist might imply a constitutive collectivity (like vortices that emerge in a zero angular momentum fluid, or poles of a magnet, it may be the case that adaptive operational closure cannot emerge in the singular). But this is an open question. Suffice it to say the empirical fact that all known forms of life are collective has clear implications for the history of organism–environment co-definition that we are examining here.

capable of caring about impending things and events in the world—is material at its core, as one would expect from a non-dualistic philosophy of nature.

The processes of mutual co-definition between organism and environment, or mutual projection between subject and world, acquire true historical power in their collective dimension. Organisms self-differentiate and produce shared worlds through common paths of interaction. The collective potentiates the structuring powers but also amplifies contradictions like the primordial tension of life, which is manifested as a primordial tension of participatory sense-making (Cuffari et al. 2015; Di Paolo et al. 2018). Life and mind never fully lose their constitutive spontaneity due to the inherent need to always keep active. This is the second enactive emphasis—one that deserves further development—the importance to conceive of the phenomena of life and mind as plural from the start. It may turn out to be that the true protagonists of Varela's relation of co-definition are collectivities of organisms and their common environment, as much as the true protagonists of Merleau-Ponty's formulation are a community of people and their shared history.

Acknowledgments

Thanks to Tom Froese, Hanne De Jaegher, and Marek McGann for helpful comments on previous versions of this chapter. This work is supported by the project "Interidentidad: Identity in Interaction" (MINECO, FFI2014-52173-P).

References

Agmon, E., Gates, A.J., and Beer, R.D. (2015). Ontogeny and adaptivity in a model protocell. In: P. Andrews, L. Caves, R. Doursat, S. Hickinbotham, F. Polack, S. Stepney et al. (eds.), *Proceedings of the European Conference on Artificial Life*. Cambridge, MA: MIT Press, pp. 216–23.

Barandiaran, X. and Egbert, M. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life*, 20, 5–28.

Barandiaran, X., Rohde, M., and Di Paolo, E.A. (2009) Defining agency: individuality, normativity, asymmetry and spatio-temporality in action. *Adaptive Behavior*, 17, 367–86.

Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/f noise. *Physical Review Letters,* 59, 381–4.

Bichat, X. (1800). *Recherche physiologiques sur la vie et la mort*. Paris: Brosson, Gabon & Cie.

Buhrmann, T. and Di Paolo, E.A. (2014). Spinal circuits can accommodate interaction torques during multijoint limb movements. *Frontiers in Computational Neuroscience*, 8, 144.

Buhrmann, T. and Di Paolo, E.A. (2017). The sense of agency—a phenomenological consequence of enacting sensorimotor schemes, *Phenomenology and the Cognitive Sciences*, 16, 207–36.

Buhrmann T., Di Paolo E.A., and Barandiaran X. (2013). A dynamical systems account of sensorimotor contingencies. *Frontiers in Psychology*, 4, 285.

Canguilhem, G. (1965/2008). *Knowledge of life*. New York: Fordham University Press.

Cuffari, E., Di Paolo, E.A., and De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences*, 14, 1089–1125.

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience*, 7, 15.

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

De Jaegher, H., Di Paolo, E.A., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14, 441–7.

Dewey, J. (1938/1991). *Logic: the theory of inquiry*. Carbonsdale and Edwardsville: Southern Illinois University Press.

Di Paolo, E.A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4, 429–52.

Di Paolo, E.A. (2009). Extended life. *Topoi*, 28, 9–21.

Di Paolo, E.A. (2015). Interactive time-travel: on the intersubjective retro-modulation of intentions. *Journal of Consciousness Studies*, 22, 49–74.

Di Paolo E.A., Barandiaran X.E., Beaton M., and Buhrmann T. (2014). Learning to perceive in the sensorimotor approach: Piaget's theory of equilibration interpreted dynamically. *Frontiers in Human Neuroscience*, 8, 551.

Di Paolo, E.A., Buhrmann, T., and Barandiaran, X.E. (2017). *Sensorimotor life: an enactive proposal*. Oxford: Oxford University Press.

Di Paolo, E.A., Cuffari, E., and De Jaegher, H. (2018). *Linguistic bodies: the continuity between life and language.* Cambridge, MA: MIT Press.

Di Paolo, E.A. and De Jaegher, H. (2012). The interactive brain hypothesis. *Frontiers in Human Neuroscience*, 6, 163.

Di Paolo, E.A., Rohde, M., and De Jaegher, H. (2010). Horizons for the enactive mind: values, social interaction and play. In: J. Stewart, O. Gapenne, and E.A. Di Paolo (eds.), *Enaction: toward a new paradigm of cognitive science*. Cambridge, MA: MIT Press, pp. 33–87.

Di Paolo, E.A. and Thompson, E. (2014). The enactive approach. In: L. Shapiro (ed.), *The Routledge handbook of embodied cognition*. London: Routledge, pp. 68–78.

Egbert, M., Barandiaran, X., and Di Paolo, E.A. (2010). A minimal model of metabolism-based chemotaxis. *PLoS Computational Biology*, 6(12), e1001004.

Egbert, M., Barandiaran, X., and Di Paolo, E.A. (2012). Behavioral metabolution: the adaptive and evolutionary potential of metabolism-based chemotaxis. *Artificial Life*, 18, 1–25.

Froese, T. (2014). Steps toward an enactive account of synesthesia. *Cognitive Neuroscience*, 5, 126–7.

Froese, T., Lenay, C., and Ikegami, T. (2012). Imitation by social interaction? Analysis of a minimal agent-based model of the correspondence problem. *Frontiers in Human Neuroscience*, 6, 202.

Froese, T. and Stewart, J. (2010). Life after Ashby: ultrastability and the autopoietic foundations of biological individuality. *Cybernetics and Human Knowing*, 17, 7–50.

Froese, T., Virgo, N., and Ikegami, T. (2014). Motility at the origin of life: its characterization and a model. *Artificial Life*, 20, 55–76.

Froese, T., Woodward, A., and Ikegami, T. (2013). Turing instabilities in biology, culture, and consciousness? On the enactive origins of symbolic material culture. *Adaptive Behavior*, 21, 199–214.

Fuchs, T. (2011). The brain—a mediating organ. *Journal of Consciousness Studies*, 18, 196–221.

Fuchs, T. (2017). *Ecology of the brain: the phenomenology and biology of the embodied mind.* New York: Oxford University Press.

Gallagher, S. (2015). The problem with 3-year olds. *Journal of Consciousness Studies*, 22, 160–82.

Gallagher, S. (2017). *Enactivist interventions: rethinking the mind.* New York: Oxford University Press.

Gallagher, S., Hutto, D., Slaby, J., and Cole, J. (2013). The brain as part of an enactive system. *Behavioral and Brain Sciences*, 36, 421–2.

Gislén, A., Dacke, M., Kröger, R.H.H., Abrahamson, M. Nilsson, D.-E., and Warrant, E.J. (2003). Superior underwater vision in a human population of sea-gypsies. *Current Biology*, 13, 833–6.

Gislén, A., Warrant, E.J., Dacke, M., and Kröger, R.H.H. (2006). Visual training improves underwater vision in children. *Vision Research*, 46, 3443–50.

Grosz, E. (2011). Matter, life, and other variations. *Philosophy Today*, 55, 17–27.

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

Koonin, E.V., Makarova, K.S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology*, 55, 709–42.

Kreines, J. (2009). The logic of life: Hegel's philosophical defense of teleological explanation in biology. In: F. Beiser (ed.), *The Cambridge companion to Hegel and nineteenth-century philosophy.* Cambridge: Cambridge University Press, pp. 344–77.

Liu, D., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S. et al. (1997). Maternal care, hippocampal glucocorticoid receptor gene expression and hypothalamic-pituitary-adrenal responses to stress. *Science*, 277, 1659–62.

Malafouris, L. (2007). Before and beyond representation: towards an enactive conception of the palaeolithic image. In: C. Renfrew and I. Morley (eds.), *Image and imagination: a global history of figurative representation*. Cambridge: McDonald Institute for Archaeological Research, pp. 289–302.

Malafouris, L. (2013). *How things shape the mind: a theory of material engagement*. Cambridge, MA: MIT Press.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybernetics and 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.

McGann, M. (2014). Enacting a social ecology: radically embodied intersubjectivity. *Frontiers in Psychology*, 5, 1321.

Merleau-Ponty, M. (1945/2012). *Phenomenology of perception* (trans. D.A. Landes). London: Routledge.

Michelini, F. (2012). Hegel's notion of natural purpose. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, 133–9.

Moreno, A. (2010). Modularity and function in early prebiotic evolution. *Origins of Life and Evolution of the Biosphere*, 40, 475–7.

Nunes-Neto, N. Moreno, A., and El Hani, C.N. (2014). Function in ecology: an organizational approach. *Biology and Philosophy,* 29, 123–41.

Odling-Smee, J.F., Laland, K.N., and Feldman, M.W. (2003). *Niche construction: the neglected process in evolution*. New Jersey: Princeton University Press.

Oyama, S. (2000). *The ontogeny of information: developmental systems and evolution.* Durham: Duke University Press.

Saborido, C., Mossio, M., and Moreno, A. (2011). Biological organization and cross-generation functions. *British Journal for the Philosophy of Science*, 62, 583–606.

Shirt-Ediss, B., Ruiz-Mirazo, K., Mavelli, F., and Solé, R.V. (2014). Modelling lipid competition dynamics in heterogeneous protocell populations. *Scientific Reports*, 4, 5675.

Silver, M. and Di Paolo, E.A. (2006). Spatial effects favour the evolution of niche construction. *Theoretical Population Biology*, 70, 387–400.

Simon, H.A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106, 467–82.

Simondon, G. (2005). *L'individuation à la lumière des notions de forme et d'information*. Grenoble: Millon.

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

Thompson, E., Palacios, A., and Varela, F.J. (1992). Ways of coloring: comparative color vision as a case study for cognitive science. *Behavioral and Brain Sciences*, 15, 1–26.

Turner, J.S. (2000). *The extended organism: the physiology of animal-built structures*. Cambridge, MA: Harvard University Press.

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

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

Varela, F.J. (2000). *El Fenómeno de la Vida*. Santiago de Chile: J.C. Sáez.

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

Villalobos, M. (2013). Enactive cognitive science: revisionism or revolution? *Adaptive Behavior*, 21, 159–67.

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.