Life is precious because it is precarious: Individuality, mortality, and the problem of meaning [*]

Tom Froese [1, 2]

[1] Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
[2] Centro de Ciencias de la Complejidad
Universidad Nacional Autónoma de México
t.froese@gmail.com

Abstract

Computationalism aspires to provide a comprehensive theory of life and mind. It fails in this task because it lacks the conceptual tools to address the problem of meaning. I argue that a meaningful perspective is enacted by an individual with a potential that is intrinsic to biological existence: death. Life matters to such an individual because it must constantly create the conditions of its own existence, which is unique and irreplaceable. For that individual to actively adapt, rather than to passively disintegrate, expresses a value inherent in its way of life, which is the ultimate source of more refined forms of normativity. This response to the problem of meaning will not satisfy those searching for a functionalist or logical solution, but on this view such a solution will not be forthcoming. As an intuition pump for this alternative perspective I introduce two ancient foreign worldviews that assign a constitutive role to death. Then I trace the emergence of a similar conception of mortality from the cybernetics era to the ongoing development of enactive cognitive science. Finally, I analyze why orthodox computationalism has failed to grasp the role of mortality in this constitutive way.

1. Introduction

Computationalism tries to explain natural phenomena in terms of the concept of computation, where "a computation is a set of objects and relations within the domain of abstract entities (as described in the logical formalisms of theoretical computer science)" (Horsman, Kendon, Stepney, & Young, this volume). In this chapter I will present criticisms of attempts to develop a general theory of life and mind based on the concepts of computation and information processing, for example info-computationalism (Dodig-Crnkovic, 2014)[1]. In essence, I argue that such attempts fail to account for the meaningful perspective that we normally experience in our lives, and contend that this is because they ignore life's deathly underpinnings, its irreducible *precariousness*.

[*] This is the penultimate draft of a chapter that will appear in G. Dodig-Crnkovic and R. Giovagnoli (Eds.), *Representation and Reality in Humans, Animals and Machines*, Springer.

[1] Note that there are alternative accounts of computation that go beyond formal structure and also appeal to properties of the physical mechanisms that realize the computations (Miłkowski, 2011, 2013; Piccinini, 2015). It remains to be seen if such mechanistic accounts can sidestep my criticisms, so from now on I use computation to refer specifically to accounts that do not bottom out in non-computational mechanisms.

This argument will not be a formal solution to the problem of explaining how there can be a subjective perspective in an objective world (which would amount to finding an algorithmic solution to the perennial problems of cognitive science, e.g. the frame problem, explanatory gap, the hard problem of consciousness, etc.). The aim of this chapter is come up with a compelling alternative argument that does not rely on the concepts of computation and information processing. Instead the argument is based on assuming continuity between life and mind, such that having a mind is dependent on biological embodiment (Thompson & Varela, 2001). Specifically, I will be making two claims: First, the potential to die is constitutive of an individual life, where the individual's process of living is considered to be a process of sense-making in its most basic sense (Weber & Varela, 2002). And second, the precariousness that is intrinsic to all organismic, and therefore also of all mental, existence is the original reason why things matter to that individual being (Jonas, [1966] 2001).

Here we therefore assume at the outset that the phenomenon to be explained is that a living being is immediately presented with a meaningful world (and not by a physicist's set of objective facts about the environment, or a computer's bits of formal information). In contrast to modern views that continue the Cartesian legacy, which does not distinguish between living beings and merely mechanical objects, we assume that to live is to always be concerned with something, most fundamentally with the continuation of one's individual manner of living. Closely related, therefore, is the classic problem of biological *individuality*. For it is only when we have the concept of an individual that we can start thinking about the role that the individual's potential to transform into an irrevocable absence, its mortality, has in shaping its lived presence (Jonas, 1992).

These considerations are applicable to even the most basic forms of life, which like all organisms must continually struggle to stave off death for yet another moment[2]. In the most minimal cases meaning will be wholly determined by basic metabolic needs. But in the specific case of human existence they also take on a highly symbolic dimension, and well-defined needs are replaced by open-ended desire (Barbaras, 2010). Awareness of our own finitude has been the inspiration for some of the oldest expressions of human culture, particularly graves, and continues to provide a dramatic source of creativity, such as Shakespeare's famous soliloquy by Hamlet or Heidegger's existential phenomenology. Indeed, Jonas (1985/86) argues that the presence of tombs and other expressions of concern for the deceased is the most incontrovertible archaeological evidence of a fully developed human mind, more so than tool or image production, because it points to an incipient metaphysics that reaches beyond life as such. Death marks the ultimate limits of living and therefore also of natural sense-making, requiring of those who desire to grasp what happens after our death the capacity to make sense of non-sense by clothing it in symbolism (Cappuccio & Froese, 2014).

[2] Certain kinds of organisms, especially single-cell organisms, can be considered to be "immortal" in the sense that they are not susceptible to deleterious effects of aging and will live as long as favorable circumstances persist. But they are still mortal in the sense that matters here: they live because they actively avoid disintegration and will die if outside their viability range.

If individuality, mortality, and meaning are interconnected concepts then they must lie at the heart of cognitive science, at least given an ambition to account for our meaningful first-person perspective. However, so far computationalism has been unable to offer a coherent approach to any of these three concepts, let alone their interdependence. While the problem of meaning has been widely discussed in cognitive science and philosophy of mind, where it has given rise to several famous thought experiments that continue to provoke debate, there is no agreed upon solution in sight. The problem of individuality also occasionally receives attention, but mainly in the philosophy of biology because evolutionary theory requires the concept of a unit of selection. As we will see, the concept of an agent in cognitive science is used frequently but has no clear definition. However, the problem of mortality has received almost no attention at all. In fact, for cognitive science it may not seem like a problem at all, but rather as a contingent fact of life on earth without any philosophical relevance. Yet from the point of view I will present in this chapter the two concepts of individuality and mortality are both essential for coming up with new ways of addressing the problem of meaning.

2. The problem of meaning

Computational theories of mind already have a long history of struggling with the problem of meaning (Dreyfus, 1972), which famously expressed itself in a number of ways, such as the Chinese room argument (Searle, 1980), the frame problem (Dennett, 1984), and the symbol grounding problem (Harnad, 1990). Essentially, the problem is to understand how things can be meaningful for a system from its own perspective, and not just from the perspective of the human observer of that system. Another way of phrasing this is that we still haven't found a response to Hume's claim that we cannot derive a value from a fact.

Nevertheless, for many researchers the notion of information seems to offer an appealing solution to this conundrum, given that it has both a strictly technical definition (following Shannon's information theory derived from the concept of entropy) and a common folk psychological interpretation (i.e. that which means something or is informative for someone). Given this terminological ambiguity it is tempting to make use of a sleight of hand, whereby the same word is used but the former concept is somehow identified with or transformed into the latter concept. Let us take a look at an example taken from info-computationalism (IC):

> Information is also a generalized concept in the context of IC, and it is always agent-dependent: *information is a difference (identified in the world) that makes a difference for an agent*, to paraphrase Gregory Bateson (1972). For different types of agents, the same data input (where data are atoms of information) will result in different information. […] Hence the same world for different agents appears differently. (Dodig-Crnkovic, 2014, p. 223)

I take it that when Dodig-Crnkovic refers to the world appearing *for* the agent she means that it appears in a meaningful way given that the information makes a difference *for* the agent. Presumably this happens via information processing, i.e. by somehow transforming the external atoms of "information" into internal meaningful "information". In other words, the mind is (metaphorically?) thought of as a container located inside the agent into which informational content (i.e., "atoms of information"), originating from the external environment, can be transferred and then manipulated. But how is it possible for environmental information, such as covariance, to be turned into mental content? There seems to be no compelling response to this problem from the perspective of traditional cognitive science (Hutto & Myin, 2013). But even if the problem of content could be solved there is still an additional problem: at what point does this content become meaningful for an agent? How do we go from the fact of there objectively being a difference in the environment to the subjective event of that dissimilarity making a difference for an agent? Even Bateson (1971) could not say much more than that the agent must be "responsive" to environmental difference[3].

Dodig-Crnkovic struggles precisely with this crucial philosophical point while trying to find a definition of information that is sufficiently broad so as to include both its technical and folk psychological meanings. She adopts Hewitt's (2007) definition and attempts to integrate it with Bateson's (1972) definition:

> "Information expresses the fact that *a system is in a certain configuration that is correlated to the configuration of another system.* Any physical system may contain information about another physical system." (Hewitt, 2007: 293, my emphasis) Combining Bateson's and Hewitt's insights, on a basic level we can state: *Information is the difference in one physical system that makes a difference in another physical system*. (Dodig-Crnkovic, 2014, p. 226)

First, we see that Hewitt assumes that one system's co-variation with another system is tantamount to the one system *containing* informational content *about* the other system. However, the latter does not straightforwardly follow from the former, as has been argued extensively by Hutto and Myin (2013). Second, there is an inherent ambiguity in the notion of "making a difference." Dodig-Crnkovic appeals to a causal interpretation, whereby one physical system causes changes to happen *in* another physical system. At least this definition is more consistent because it describes both systems from the perspective of an external observer. However, leaving aside the problem of interpreting correlation as causation, the idea of cause and effect is still not enough to explain how making a difference *in* one of the systems could make a difference *for* that system.

One of the key problems is that information theory is incapable of providing a coherent definition of an individual agent. Dodig-Crnkovic adopts Hewitt's actor model of computation, and claims that "Hewitt's 'computational devices' are conceived as computational agents – informational structures capable of acting

[3] Hutto and Myin (2013) also opt for such a non-autopoietic, behavior-based approach to basic minds. However, responsiveness to environmental difference (or to covariance) is not sufficient to account for the emergence of meaning (see also Froese, 2014). Interestingly, Bateson (1970) avoided this problem because he assumed that the environment itself embodied a larger God-like Mind of which an individual mind is only a subsystem. Forms of panpsychism have also been attractive for contemporary information theoretic approaches to consciousness (Tononi, 2008; Chalmers, in press), although it has also been used as a *reductio ad absurdum* (Bishop, 2009).

on their own behalf" (Dodig-Crnkovic, 2014, p. 225). This brings us to the deep problems of defining agency, action, and even responsibility. Is a computer an agent in any relevant sense? Or a thermostat? Such loose definitions have been widely adopted in AI, but they are unsatisfactory for a number of reasons (Froese & Ziemke, 2009). We might as well ask ourselves: what is *not* an agent on that view? If a computer can be said to be acting on its own behalf, can we not say the same thing about a planet moving around the sun or about any other physical systems? Indeed, info-computationalism does not hesitate to adopt a definition of agency that applies to physical systems at all scales: "an agent can be as simple as a molecule" (Dodig-Crnkovic, 2014, p. 225). But to say that every physical system is an "agent" in some sense brings us no closer to explaining why there is a meaningful world for us and other living beings. For if every difference of any system is information, and any change in that information is computation, and any system undergoing that computation is an agent, then we have managed to unify everything in general, but at the steep cost of failing to explain anything in particular. Without a story about how the agency of living individuals, including ourselves, differs in essential aspects from the dynamics of mere objects we are forced to either side with some version of panpsychyism by elevating objects to the status of genuine individuals or embrace a form of nihilism by reducing living individuals to the status of mere objects.

3. The problem of individuality

Computationalism struggles to come up with a coherent notion of individuality, which would require it to determine the other's boundaries in a manner that would allow that individual to transcend our determination from the outside, i.e. for the individual to at least partially escape complete reduction to an observer's perspective. Both computation and information are inherently observer-relative concepts that preclude them from being intrinsic properties of the phenomena (Eden, this volume; Deacon, this volume), and are therefore unsuitable for this task. We will return to this point at the end of this section. For now it is crucial to note that this criticism of relativity should not be misunderstood with reference to an absolute reality in itself. The point is not to remove the role of the observer altogether and adopt a view from nowhere, but to make space in the relationship that the observer has with the observed to allow for that other end of their relation to at least appears as having some intrinsic properties that are self-determining.

This is why enactive cognitive science is founded on the concept of autopoiesis (Thompson, 2007), which can be loosely defined as a network of processes that form a whole because the processes are enabled by each other. In other words, this a low level concept of individuality in the form of a living system's self-organized identity, which can be realized in such a diverse and nested manner that even a single organism, including ourselves, can be thought of a "meshwork of selfless selves" (Varela, 1991). The key advantage of autopoietic theory is that it allows any living system, if we distinguish its boundaries appropriately, to appear to us as being autonomous, i.e. as spontaneously self-distinguishing. This is the first step toward a flexible and operational theory of individual agency, which additionally includes asymmetrical regulation of the autonomous system's environmental coupling in accordance with its own normativity (Barandiaran, Di Paolo, & Rohde, 2009). When starting from such a definition of agency, based on the concept of a precarious self-producing network of processes, the system's emergent behavior is a expression of its ongoing metabolic self-realization and is therefore intrinsically related to satisfying the needs that allow the individual to maintain its way of living (Froese, Virgo, & Ikegami, 2014). This inner relation between being and doing is one reason why the world that appears from the perspective of the agent makes sense to it. And this is also why a living being is always situated in a meaningful world, whereas an artificially intelligent system, whose systemic identity is completely and arbitrarily defined from the outside, has to face the problem of meaning (Di Paolo, 2010; Froese & Ziemke, 2009).

A corollary of this account of individuality is that by definition it does not permit the formulation of a complete model of any specific example of an individual, at least not as long as it can be said to be alive. This applies as much to real as to virtual organisms. In the case of a real organism, all measurement depends on an interaction between an observer and the system, and a full determination of a living system from the outside perspective would only be possible by engaging in interactions that destroy the self-determination that is autonomously enacted by that system from the inside, i.e. by killing it. In the case of a virtual organism we do not have to kill it in order to know it completely, since we have full access to the code which implements it, but the final result is the same. Having a complete simulation model of an individual running on a computer would amount to that individual not transcending our determination from the outside, and thus failing to overcome the limitation of pure observer-relativity.

To be fair, more work needs to be done on how this concept of an autonomous systemic identity, which applies even to the most basic of living systems, scales up to the individual self that is characteristic of human existence (Kyselo, 2014). And it is still not clear if even the basic concept can respond to all the challenges that have been associated with the notion of an individual in contemporary biology (Clarke, 2010). Similarly, the concept of agency requires further work. For example, the phenomenologically inspired concept of normativity, which lies at the root of the distinction between intentional action and passive movement, is not without its critics even among researchers who are otherwise sympathetic to an enactive approach (Barrett, in press; Villalobos & Ward, 2015). I mention these issues here to highlight that the enactive theory of individual agency is far from complete and is an ongoing project.

At the same time it must be acknowledged that research on computational and information theoretic measures of aspects of biological and mental organization, for example emergence, self-organization, homeostasis, autopoiesis, and even consciousness, continues to advance (e.g., Fernández, Maldonado, & Gershenson, 2014; Oizumi, Albantakis, & Tononi, 2014). But so far these measures are limited by the lack of a coherent concept of individual agency. And there are compelling reasons to think that they cannot account for this individuality even in principle due to their reliance on principles that are inherently observer relative.

Information theory can only account for a system's information from the perspective of the external observer of that system, but this says nothing about any putative intrinsic perspective enjoyed by that system for itself (Beaton & Aleksander, 2012). Moreover, this dependence on the external observer entails that the reference (i.e. the "aboutness") of information is not an intrinsic property of the information-bearing medium, either (Deacon, this volume). Related worries about observer-relativity also apply to computational accounts, which suffer from a reliance on interpretation from the outside to determine the specific form of the computational process and its particular meaning (Bishop, 2009; Eden, this volume).

I agree that this reliance on observer-relative notions is problematic, but it is not fatal since computationalism can follow the enactive approach in adopting the presence of autopoietic self-distinction as the mark of an autonomous individual. Info-computationalism, for example, cites the work by Maturana and Varela on autopoiesis as one important influence (Dodig-Crnkovic & von Haugewitz, this volume). However, a simulation model of autopoiesis is not sufficient for genuine individuality given that a model allows complete external determination.

4. The problem of mortality

The more serious problem that I am concerned with is thus computationalism's abstraction from the concreteness of biological existence, which prevents it from grasping the precariousness of an irreplaceable living being. Information theory is one way of formalizing this abstraction. I therefore agree with the assessment of Gershenson (2014, p. 241), who argues that "considering only information, one cannot distinguish the physical from the virtual" and that to have a complete scientific account "it is not enough to consider only the organization/information of systems; their substrate and their relation must also be considered."

To be fair, it is true that processes involving information and computation are necessarily also dependent on some physical implementation that realizes them (Dodig-Crnkovic & von Haugewitz, this volume). But the concrete materiality of their implementation does not necessarily shape the form of these processes, which can after all be multiply realizable and substrate independent. Given some specification of a computation such as OR(True,False) it does not matter whether it happens to be realized as physical processes, or executed by a virtual machine on my laptop, or by my use of pen and paper, or in my imagination. At the level of computation the process of a logical OR is identical across all cases.

Given that computationalism does not distinguish the physical from the virtual it comes as no surprise that death, as the irrevocable disintegration of autonomous individuality, has hardly ever been problematized from an information theoretic perspective (although there is an exception, Gershenson, 2011, to which we will return later on). In brief, virtual agents are immortal because their existence is fully exhausted by informational structures that can be indefinitely recreated in an absolutely identical manner. Death is therefore relegated to an unfortunate fact of life on earth that could conceivably be avoided under other circumstances, such as with more advanced medicine and technology.

To be sure, computationalism is certainly not alone in this neglect of death. Apart from mortality's role in population statistical considerations and the principles of evolutionary biology (Sterelny & Griffiths, 1999), it is a marginalized topic in the mainstream sciences of life and mind in general. Various reasons can be offered for this neglect, both cultural and theoretical. Bateson (1970) relates it to modernity's rejection of all religious narrative which nevertheless lingers as a culture steeped in mind-body dualism: "It is understandable that, in a civilization which separates mind from body, we should either try to forget death or to make mythologies about the survival of transcendent mind." Indeed, death has been abstracted away by computationalism as irrelevant to understanding the basic principles of the mind. For example, even the most realistic simulations of the brain to date basically treat the neural network is if it were as immortal as the mathematical equations that model its activity. No relevance is seen in treating the brain as an organ of a precarious body in need of metabolic and dynamic self-renewal and thus forced to be an open system in interaction with the world.

It is not my intention to provide a more detailed analysis of the reasons for this scientific neglect of death here (but see Varela, 1997, pp. 131-136). Instead I only indicate that the current scientific perspective is rather unusual when compared to many traditional worldviews, which assign a constitutive role to mortality in their representations of reality. I will consider two such worldviews in order to help us to bracket our modern attitudes toward the uselessness of death. Then I will return to the scientific perspective and highlight some defining moments in the history of systems biology and enactive cognitive science, which reveal that some of these aspects of traditional worldviews are currently being recovered, in particular that mortality plays a constitutive role in an individual's life. Finally, I conclude with an analysis of the limitations of computationalism with respect to coming to terms with this kind of perspective on mortality.

5. The role of death in traditional worldviews

Mortality may seem quite useless to us today in our youth obsessed culture, but this stance is not universally shared across cultures. Death my also be considered to play an essential role in life. I will briefly illustrate this alternative perspective with two examples of foreign ancient cultures.

The family of cultures present throughout ancient Mesoamerica recognized that there was a circular interdependence between life and death. We can see this clearly in people's relationship with maize. Maize cultivation was a necessary condition for the rise of civilizations in this area, primarily because it allowed populations to expand to sufficiently large numbers (Coe & Koontz, 2013). Yet people were a necessary condition for the survival of maize, too. It required human help to free the seeds from the ears of corn, for otherwise they have problems germinating and fail to spread sufficiently. So while humans are the maize plant's principal cause of death (via harvest), humans are also an essential condition of its long-term survival and flourishing (via sowing). This unification of the duality of life and death into a circular, dynamically integrating whole was culturally manifested in a variety of ways.

This Mesoamerican relationship between life and death is often described as a form of duality, although it must be emphasized that no independence of the two terms is implied. It is a duality that recognizes the essential interdependence of opposites and thus implies complementarity. It can be traced as far back as the early formative period of central Mexico, from which ceramic masks of a half living, half skeletal face have been uncovered (Miller & Taube, 1993, p. 81).

The use of such dualistic pairings is one of the basic principles of Mesoamerican thought. The interaction between the two halves of a duality was what we would call nonlinear, in that something new would emerge from their coupling. For example, in Nahuatl ritual speech the phrase "fire-water" signified war. The principle of duality was so important that it was deified as *Ometeotl* (the "two god") and assigned to the highest level of heaven, *Omeyocan* ("place of duality") in the form of a couple, *Ometecuhtli* and his consort *Omecihuatl*. The Aztecs venerated *Ometeotl* as the supreme creative principle, a self-generating being, in which male and female principles were joined. These principles in turn belonged to a larger group of oppositions where, for example, one side would include male, life, and day, while the other side would include female, death, and night (Miller & Taube, 1993, p. 81). This suggests that the creative principle of self-generation is itself also co-constituted as one complex unity by the interdependence of the specific principles of regeneration (life) and decay (death).

In Hinduism we find that time is circular (like it was in ancient Mesoamerica), and that each cycle of cosmic time, known as a *kalpa*, features a tripartite pattern of maintenance, creation and destruction that is enacted by the *trimurti* of gods (Coe, 2003, p. 82). Vishnu, Brahma, and Shiva all had relatively distinct roles, namely to preserve, to create, and to destroy the world, respectively. The unity of the *trimurti* can be seen in one of Hinduism's most important objects of worship, the male *linga*, a symbolic phallus: the top of the *linga* symbolizes Shiva, the middle Brahma, and the base Vishnu. Interestingly, the most important god of the three is Shiva, who is also symbolized by the *linga* as a whole.

We see the principle of destruction at work as well in the idea of *samsara*, the circle of conditioned human existence. The organization of the cosmic cycles is matched by a belief in reincarnation, the cycle of personal life and death. In fact, according to Buddhist philosophy, we can even find it within the timescale of our lives: it is inherent in every moment of our existence. What we experience as the constant present is actually dynamically maintained as a "circle of arising and decay of experience [that] turns continuously" (Varela, Thompson, & Rosch, 1991, p. 80). The concept of death plays a key role in this process:

> Whenever there is birth, there is death; in any process of arising, dissolution is inevitable. Moments die, situations die, and lives end. Even more obvious than the uneasiness of birth is the suffering (and lamentation, as is said) experienced when situations or bodies grow old, decay, and die. In this circular chain of causality, death is the causal link to the next cycle of the chain. The death of one moment of experience is,

> within the Buddhist analysis of causality, actually a causal precondition for the arising of the next moment. (Varela, et al., 1991, p. 115)

What this quotation makes clear is that a generalized concept of death can be taken as one of existence's essential principles. I also note that this quotation was taken from the foundational text of the enactive approach to cognitive science, namely *The Embodied Mind: Cognitive Science and Human Experience*. This was the first but not the only way in which the enactive approach begun to recognize that death is an essential explanatory principle for its theory of life and mind. Let us take a closer look at its history.

6. The role of death in enactive cognitive science

Enactivism is the latest installment in a long line of intellectual movements that built on each other and were equally shunned by mainstream thinkers. Many important ideas and methods of enactive theory can be traced back to the early cybernetics era and especially to that era's end in the work of Ashby, and to that work's later expansion and refinement in Maturana and Varela's biology of cognition, until we finally arrive at the first formulations of an enactive cognitive science (Froese, 2010, 2011; Froese & Stewart, 2010). Each of these phases contributed essential insights, some of which I want to highlight.

One of the key insights of cybernetics was that it is possible to devise a systems theory of *self-maintenance* based on the principle of negative feedback. Famous examples include the Watt's governor and the thermostat. Maturana and Varela's ([1973] 1980) theory of autopoiesis built on these insights, and added the crucial insight that living systems are not only self-maintaining, but also *self-producing*, which distinguishes them from AI and robotics (Froese & Ziemke, 2009).

In these two stages of conceptual development a positive role of death starts to be prefigured. Ashby (1947) founded his ultrastability theory on the idea of a system *breaking* and thereby losing its original systemic identity (due to changes of its parameters) in the process. Maturana and Varela recognized that *decay* is an indispensable property of the components for the formation of the autopoietic system. However, ultimately neither the breaking nor the decay were conceived of as somehow affecting the identity of the whole system.

Ashby's (1960) homeostat was built precisely so as to remain a homeostat, even while specific parts of it were occasionally "breaking" (undergoing random parameter changes). Note that the homeostat's systemic identity and the changes it could undergo were pre-defined externally by Ashby, and this lack of autonomy precludes a genuine role of precariousness (Froese, 2013). And Maturana and Varela concluded, in line with the abstractness of cybernetics and general systems theory, that the "properties of the components of an autopoietic system *do not* determine its properties as a unity" (Varela, Maturana, & Uribe, 1974, p. 192), thereby banishing the effects of decay from the domain of the system as a unity. I agree with Di Paolo (2009) that Maturana's doctrine of non-intersecting domains, although a well-intentioned reaction against physicalist reductionism, has had the unfortunate side-effect of preserving mind-body dualism in another format. Following Bateson, it is therefore not surprising that death was once again neglected. Importantly, this doctrine prevents decay, as an inherent property of the chemical components, to be meaningfully related to the mortal existence of the living, as if the instability of the components had nothing to do with the precariousness of the whole.[4]

Varela later overcame this mere contingency of the principles of breaking and disintegration in two important ways. He came to see the breakdowns of animal behavior and human experience as the "birthplace of the concrete" in which the cognitive agent and their immediate world becomes spontaneously reconstituted and creatively rearticulated in an action-appropriate manner (Varela, 1995). To illustrate this idea, Varela asks us to imagine what happens when we reach for our wallet and realize that it is no longer there – after a transitory moment of confusion we will find ourselves in a new task-specific being-in-the-world geared toward the rapid recovery of our wallet. We can see how this idea of breakdowns as the birthplace of the concrete resembles the Buddhist idea of the death of one moment causing a new one to emerge.

More generally, Varela also acknowledged the essential role of mortality in his later enactive theory of the organism, following a close reading of Jonas' ([1966] 2001) phenomenological philosophy of life (Weber & Varela, 2002). And he was also forced to recognize its importance personally as he dealt with the rapidly deteriorating state of his own body toward the end of his life. He concluded the phenomenological analysis of his harrowing experiences of undergoing organ transplantation with the poignant statement: "Somewhere we need to give death back its rights" (Varela, 2001, p. 271). This intertwining between third-person scientific theory and first-person existential insight is a general characteristic of enactivism, for example as practiced in neurophenomenology (Bitbol, 2002), but it is certainly at its most demanding and intimate when the phenomenon under consideration is death.

For Jonas ([1966] 2001) life and death are two sides of the same coin, and out of this complementary unity arises something novel: individual beings who are concerned about maintaining their own form of being. By constructing their own boundaries under far-from-equilibrium conditions, living beings determine their own individuality and their relationship with the world, and they do so in a way that gives them intrinsic value. Starting from the phenomenological insight that we know ourselves to be more than pure mechanisms devoid of a meaningful perspective, and accepting this as a fundamental fact needing to be explained, he set out to argue for the essential role of mortality in accounting for that meaning:

> with metabolizing existence not-being made its appearance in the world as an alternative embodied in the existence itself […]: intrinsically qualified by the threat of its negative it must affirm itself, and existence affirmed is existence as a *concern*. Being has become a task rather than a

---

[4] In more recent formulations of autopoietic theory, Maturana (2011) has started to emphasize that autopoietic systems are a kind of molecular system. But more needs to be done to unpack the implications of this restriction to the chemical domain in terms of our understanding of the phenomenon of life, implications that enactive theory is unfolding (Froese & Stewart, 2012).

> given state, a possibility ever to be realized anew in opposition to its ever-present contrary, not-being, which inevitably will engulf it in the end. […] Are we then, perhaps, allowed to say that mortality is the narrow gate through which alone *value* […] could enter the otherwise in-different universe? […] Only in confrontation with ever-possible not-being could being come to feel itself, affirm itself, make itself its own purpose. (Jonas, 1992, pp. 35-36)

Note that this is not a causal relationship between death and life. The upshot is that we cannot separate an organism's systemic identity from its precarious material realization without losing the capacity to account for the meaning and the intrinsic teleology of life we all know from our personal experience. Death is dependent on a certain material configuration: "Because form that desires itself in a purposeful manner is happening only in matter to which form is not its entropically 'natural' state, there is always the possibility, and final certainty, of death" (Weber & Varela, 2002, p. 119). On this view, a meaningful perspective and mortality are inextricably linked in their material embodiment.

Accordingly, we seem to find in recent enactivism something akin to the ancient Mesoamerican principle of complementarity, *ometeotl*, as applicable to death and life. And in its historical development there were three principles of particular significance that resemble the elements of the Hindu *trimurti*: self-maintenance (Ashby's cybernetics), self-production (Maturana's autopoietic theory), as well as death and precariousness (Varela's enactivism). This convergence to similar principles of human existence, under such hugely diverse circumstances, makes sense if we consider that all major worldviews are shaped by universally shared aspects of human existence. Moreover, enactivism was also inaugurated with the explicit aim of incorporating phenomenological invariants of human experience into cognitive science. In other words, perhaps this is a case of intersubjective validation at the intercultural level.

Yet this rediscovery of precariousness as an explanatory principle only occurred after around half a century of cognitive science and only on the sidelines of the mainstream. Is the problem of mortality in a blindspot of computationalism? If we can better understand this neglect we will get a better grasp of the limitations of the computational theory of mind as an account of human existence.

7. On the impossibility of a virtual death

From the perspective of computationalism it is a purely contingent fact about a physical computational system that it can be destroyed, for example by smashing it to bits and pieces. But this *potential* destructibility of the implementation at the physical level is completely irrelevant for the functions that these systems are abstractly realizing at the computational level[5]. In this respect I disagree with

---

[5] It might be argued that the biological phenomenon known as programmed cell death, whereby cells within a multicellular organism spontaneously disintegrate, is an exception: their lifespan is related to properties of their chromosomes and this relationship can be analyzed in information theoretic terms (Zenil, Schmidt, & Tegnér, in press). Moreover, such disintegration of cells plays a variety of functions in a multicellular organism. Clearly, information theory can therefore help in

Gershenson's (2011) analysis of death, in which no ontological distinction is drawn between the disintegration of a real individual and of a virtual "agent":

> If we can create again a living system with the same organization, did it die in the first place? I think the answer should be in the affirmative. The fact that an organism – artificial or natural – can easily be replaced or regenerated does not mean that the particular instantiation of its organization is not lost. (Gershenson, 2011, p. 3)

In the case of a real organism an organizational perspective on life, whereby the identity of a living system is defined only by its organization, makes it tempting to assume that death can be cheated by re-creating that same organization at a later point in time. Early conceptions of autopoietic theory can be criticized of promoting such an abstract stance in which the organization was considered independently from its material realization (Froese & Stewart, 2010). However, to identify the actual living being with its description as a living system simply confuses the description of its organization with the reality of its being. The description can never exhaust the actual reality because describing a physical phenomenon's organization depends on an act of abstraction that by definition distinguishes the abstract organization from the concrete materiality. A better way to think about the relationship between a real individual organism and its systemic organization is in terms of instantiation of a species. A particular instantiation is irreplaceable, on this point I agree with Gershenson, although it would have been more precise for him to say that it is the species, or at least a category, of organism that could be easily replaceable.

However, in the case of a virtual agent the reality does not exceed its abstract organization; the two are one and the same because in the computational realm there are nothing but abstract entities in the first place. In this case Gershenson's argument about the death of distinct material instantiations no longer applies because the computational level of the agent can be implemented so as to be formally independent from the underlying material substrate. In that case there is nothing more to a virtual agent but its organization at any point in time, and thus at the level of computation nothing could in principle distinguish it from a later, independent instantiation of that very same identity. It is as if there is only one virtual agent defined by its purely logical, and therefore immortal, identity, which can be realized again and again in the form of indistinguishable clones.[6]

There is therefore a crucial difference between the existence of a real material organism and the persistence of a virtual agent in a computer simulation. Only the latter can return from disintegration to exactly its former being as if nothing had happened. Given that the virtual agent's identity is completely exhausted by its formal organization, that identity remains what it is even if it is not currently

analyzing some of the causes and functions of cellular death. But here death is approached as a contingent fact of life on earth rather than as something essential to it.

[6] Incidentally, this is why no matter how hard copyright enforcement authorities try to convince people that the unpermitted replication of digital products is the same as stealing, there will always be an essential difference between stealing a car (or a handbag, a television, a movie – all distinct physical items) and downloading copies of movies (all indistinct virtual clones).

realized. It's state of death, or not-being, is only relative to the end of a particular instantiation from which it ultimately remains independent. Conversely, the real organism is a unique and irreplaceable individual, its future horizon necessarily limited in principle by the inevitable possibility of irrevocable death.

Following this contrast between the death of a real organism and the ending of an instantiation of a virtual agent we must refine Jonas' concept of mortality to deal with the technological advances of our times. Virtual agents can disintegrate but this is not sufficient for considering them to exist precariously. It is not just the potential for disintegration that is essential, but also the fact that this event is irrevocable once it occurs.[7] An eternal non-being rather than just a temporarily non-realized being must follow death. In other words, in order to give rise to a meaningful perspective the precariousness of an individual cannot be separated from its uniqueness. Only a real organism in its continual struggle to continue its way of life can therefore be genuinely concerned with its own existence, with the world, and with the lives of others. It may be strange to consider the potential for irrevocable non-being as constitutive of meaning, but this kind of change in our explanatory framework may be exactly what we need in order to explain how mind emerged from matter (Deacon, 2012).

8. Conclusions

Since the beginnings of the cybernetics era over half a century ago, system-based approaches to life and mind have been recovering the essential role of death in the modern scientific worldview. From the cybernetics of self-maintenance, to the biology of cognition of self-production, to contemporary enactive theory of precariousness, we rediscovered the same interlinked principles that have been at the core of important traditional worldviews for millennia. I believe this is a good thing, for it indicates that cognitive science is once again becoming aligned with human experience. This gives hope that the general crisis of the sciences, which Husserl diagnosed in the first half of the last century, is coming to an end. We are finally returning to the concrete domain from which all scientific activity must start in the first place, our pre-theoretical lifeworld.

Yet enactive theory is not going to be welcomed by the majority of researchers because it implies uncomfortable rethinking of basic assumptions, but because it cannot be separated from our personal ideas about life and death. For example, it implies that popular ideas about how to make people immortal by turning them into purely virtual selves are misguided. While those ideas may appear to be life affirming, they are actually stripping life of its essential nature – its precarious and therefore meaningful existence. Conversely, taking seriously the biologically embodied mind cannot avoid bringing us face to face with the inevitability of our own finitude, which conflicts with the transhumanist goal of defeating death by engineering our bodies to stay forever young (e.g., Young, 2006). This conflict does not have to be situated at the ontological level, since even living bodies that stop aging and never get sick would still be mortal. And human existence can be

[7] Di Paolo (2009) highlights that this constitutive role of precariousness marks a break with the tradition of functionalism: death is not a function, which could be reverted, but it is rather the cessation of all function.

precarious in more respects than just at the basic biological level; it also has a lot to do with the continuation of a way of life rather than just with life itself. But there is a mismatch at the conceptual level: transhumanism views mortality as a burden to be removed or at least as something to be postponed indefinitely by scientific progress, rather than as constitutive of a meaningful way of life.

We may speculate that a human being who is leading a way of life without any real sense of finitude will face serious existential issues in the long run. And the alternative is actually not as bad as it may seem. For as Jonas, following his mentor Heidegger, emphasized: facing up to our own inevitable death is only a burden as long as we ignore mortality's role in making our life meaningful in the first place. Moreover, as conscious beings we enjoy the additional privilege of being able to take advantage of this insight into our finitude in order to realize the full potential of our lives with the awareness that each moment is as precious as it is precarious.

> As to our mortal condition as such, our understanding can have no quarrel about it with creation unless life itself is denied. As to each of us, the knowledge that we are here but briefly and a nonnegotiable limit is set to our expected time may even be necessary as the incentive to number our days and make them count. (Jonas, 1992, p. 40)

Acknowledgements

Ezequiel Di Paolo provided constructive feedback on an earlier version of this manuscript. I thank the many reviewers whose detailed comments and criticisms helped to substantially sharpen the final version.

References

Ashby, W. R. (1947). The nervous system as physical machine: With special reference to the origin of adaptive behavior. *Mind, 56*(221), 44-59.

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

Barandiaran, X., Di Paolo, E. A., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive Behavior, 17*(5), 367-386.

Barbaras, R. (2010). Life and exteriority: the problem of metabolism. In J. Stewart, O. Gapenne & E. A. Di Paolo (Eds.), *Enaction: Toward a New Paradigm for Cognitive Science* (pp. 89-122). Cambridge, MA: The MIT Press.

Barrett, N. F. (in press). The normative turn in enactive theory: An examination of its roots and implications. *Topoi*.

Bateson, G. (1970). Form, substance, and difference. *General Semantics Bulletin, 37*, 221-245.

Bateson, G. (1971). The cybernetics of self: A theory of alcoholism. *Psychiatry, 34*(1), 1-18.

Bateson, G. (1972). *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. New York: Ballentine Books.

Beaton, M., & Aleksander, I. (2012). World-related integrated information: Enactivist and phenomenal perspectives. *International Journal of Machine Consciousness, 4*(2), 439-455.

Bishop, J. M. (2009). A Cognitive Computation Fallacy? Cognition, Computations and Panpsychism. *Cognitive Computation, 1*, 221-233.

Bitbol, M. (2002). Science as if situation mattered. *Phenomenology and the Cognitive Sciences, 1*, 181-224.

Cappuccio, M., & Froese, T. (2014). Introduction. In M. Cappuccio & T. Froese (Eds.), *Enactive Cognition at the Edge of Sense-Making: Making Sense of Non-Sense* (pp. 1-33). Basingstoke: Palgrave Macmillan.

Chalmers, D. J. (in press). Panpsychism and panprotopsychism. In T. Alter & Y. Nagasawa (Eds.), *Consciousness in the Physical World: Perspectives on Russellian Monism*. New York: Oxford University Press.

Clarke, E. (2010). The problem of biological individuality. *Biological Theory, 5*(4), 312-325.

Coe, M. D. (2003). *Angkor and the Khmer Civilization*. London, UK: Thames & Hudson.

Coe, M. D., & Koontz, R. (2013). *Mexico: From the Olmecs to the Aztecs*. London, UK: Thames & Hudson.

Deacon, T. W. (2012). *Incomplete Nature: How Mind Emerged from Matter*. New York, NY: W. W. Norton & Company.

Deacon, T. W. (this volume). Information and reference. In G. Dodig-Crnkovic & R. Giovagnoli (Eds.), *Representation and Reality: Humans, Animals and Machines*: Springer.

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, Machines and Evolution: Philosophical Studies* (pp. 129-152). Cambridge: Cambridge University Press.

Di Paolo, E. A. (2009). Extended life. *Topoi, 28*(1), 9-21.

Di Paolo, E. A. (2010). Robotics inspired in the organism. *Intellectica, 1-2*(53-54), 129-162.

Dodig-Crnkovic, G. (2014). Info-computational constructivism and cognition. *Constructivist Foundations, 9*(2), 223-231.

Dodig-Crnkovic, G., & von Haugewitz, R. (this volume). Reality construction in cognitive agents through processes of info-computation. In G. Dodig-Crnkovic & R. Giovagnoli (Eds.), *Representation and Reality: Humans, Animals and Machines*: Springer.

Dreyfus, H. L. (1972). *What Computers Can't Do: A Critique of Artificial Reason*. New York, NY: Harper and Row.

Eden, Y. (this volume). Being 'simple-minded': models, maps and metaphors and why the brain is not a computer. In G. Dodig-Crnkovic & R. Giovagnoli (Eds.), *Representation and Reality: Humans, Animals and Machines*: Springer.

Fernández, N., Maldonado, C., & Gershenson, C. (2014). Information measures of complexity, emergence, self-organization, homeostasis, and autopoiesis. In M. Prokopenko (Ed.), *Guided Self-Organization: Inception* (pp. 19-51). Berlin: Springer-Verlag.

Froese, T. (2010). From cybernetics to second-order cybernetics: A comparative analysis of their central ideas. *Constructivist Foundations, 5*(2), 75-85.

Froese, T. (2011). From second-order cybernetics to enactive cognitive science: Varela's turn from epistemology to phenomenology. *Systems Research and Behavioral Science, 28*, 631-645.

Froese, T. (2013). Ashby's passive contingent machines are not alive: Living beings are actively goal-directed. *Constructivist Foundations, 9*(1), 108-109.

Froese, T. (2014). Radicalizing Enactivism: Basic Minds without Content. Daniel D. Hutto and Erik Myin. Cambridge, Massachusetts: MIT Press, 2013, 206 pages, $35.00 hardcover. *The Journal of Mind and Behavior, 35*(1-2), 71-82.

Froese, T., & Stewart, J. (2010). Life after Ashby: Ultrastability and the autopoietic foundations of biological individuality. *Cybernetics & Human Knowing, 17*(4), 83-106.

Froese, T., & Stewart, J. (2012). Enactive cognitive science and biology of cognition: A response to Humberto Maturana. *Cybernetics & Human Knowing, 19*(4), 61-74.

Froese, T., Virgo, N., & Ikegami, T. (2014). Motility at the origin of life: Its characterization and a model. *Artificial Life, 20*(1), 55-76.

Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence, 173*(3-4), 366-500.

Gershenson, C. (2011). What does artificial life tell us about death? *International Journal of Artificial Life Research, 2*(3), 1-5.

Gershenson, C. (2014). Info-computationalism or materialism? Neither and both. *Constructivist Foundations, 9*(2), 241-242.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*, 335-346.

Hewitt, C. (2007). What is commitment? Physical, organizational, and social. In P. Noriega, J. Vazquez-Salceda, G. Boella, O. Boissier & V. Dign (Eds.), *Coordination, Organizations, Institutions, and Norms in Agent Systems II* (pp. 293-307). Berlin: Springer.

Horsman, D. C., Kendon, V., Stepney, S., & Young, J. P. W. (this volume). Abstraction and representation in living organisms: when does a biological system compute? In G. Dodig-Crnkovic & R. Giovagnoli (Eds.), *Representation and Reality: Humans, Animals and Machines*: Springer.

Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: The MIT Press.

Jonas, H. (1985/86). Werkzeug, Bild und Grab: Vom Transanimalischen im Menschen. *Scheidewege, 15*, 47-58.

Jonas, H. (1992). The burden and blessing of mortality. *The Hastings Center Report, 22*(1), 34-40.

Jonas, H. ([1966] 2001). *The Phenomenon of Life: Toward a Philosophical Biology*. Evanston, IL: Northwestern University Press.

Kyselo, M. (2014). The body social: An enactive approach to the self. *Frontiers in Psychology, 5*(986). doi: 10.3389/fpsyg.2014.00986

Maturana, H. R. (2011). Ultrastability ... autopoiesis? Reflexive response to Tom Froese and John Stewart. *Cybernetics & Human Knowing, 18*(1-2), 143-152.

Maturana, H. R., & Varela, F. J. ([1973] 1980). Autopoiesis: The organization of the living *Autopoiesis and Cognition: The Realization of the Living* (pp. 59-140). Dordrecht: Kluwer Academic.

Miłkowski, M. (2011). Beyond formal structure: A mechanistic perspective on computation and implementation. *Journal of Cognitive Science, 12*, 359-379.

Miłkowski, M. (2013). *Explaining the Computational Mind*. Cambridge, MA: MIT Press.

Miller, M., & Taube, K. (1993). *An Illustrated Dictionary of the Gods and Symbols of Ancient Mexico and the Maya*. London, UK: Thames & Hudson.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology, 10*(5), e1003588. doi: 10.1371/journal.pcbi.1003588

Piccinini, G. (2015). Computation in physical systems. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Summer 2015 Edition)* (URL = http://plato.stanford.edu/archives/sum2015/entries/computation-physicalsystems/).

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417-424.

Sterelny, K., & Griffiths, P. E. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: The University of Chicago Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418-425.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin, 215*, 216-242.

Varela, F. J. (1991). Organism: A meshwork of selfless selves. In A. I. Tauber (Ed.), *Organism and the Origins of Self* (pp. 79-107). Dordrecht, Netherlands: Kluwer Academic Publishers.

Varela, F. J. (1995). The re-enchantment of the concrete: Some biological ingredients for a nouvelle cognitive science. In L. Steels & R. Brooks (Eds.), *The Artificial Life Route to Artificial Intelligence* (pp. 11-22). Hove, UK: Lawrence Erlbaum Associates.

Varela, F. J. (2001). Intimate distances: Fragments for a phenomenology of organ transplantation. *Journal of Consciousness Studies, 8*(5-7), 259-271.

Varela, F. J. (Ed.). (1997). *Sleeping, Dreaming and Dying: An Exploration of Consciousness with the Dalai Lama*. Boston: Wisdom Publications.

Varela, F. J., Maturana, H. R., & Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems, 5*, 187-196.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Villalobos, M., & Ward, D. (2015). Living systems: Autonomy, autopoiesis and enaction. *Philosophy & Technology, 28*(2), 225-239.

Weber, A., & Varela, F. J. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences, 1*, 97-125.

Young, S. (2006). *Designer Evolution: A Transhumanist Manifesto*. Amherst, NY: Prometheus Books.

Zenil, H., Schmidt, A., & Tegnér, J. (in press). Causality, information and biological computation: An algorithmic software approach to life, disease and the immune system. In S. I. Walker,

P. C. W. Davies & G. Ellis (Eds.), *Information and Causality: From Matter to Life*. Cambridge: Cambridge University Press.