Varela's Legacy for ALIFE: from Enactive to Enlightened AI

Elena Antonova[1,2]

[1]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London
[2]Adaptive Systems Research Group and Royal Society Wolfson Biocomputation Research Laboratory
Centre for Computer Science and Informatics Research, University of Hertfordshire
elena.antonova@kcl.ac.uk

Keynote Abstract

Francisco Varela started his scientific pursuits in the field of biology migrating to cognitive neuroscience later in his research career. To understand Varela's legacy, we need to trace the evolution of his thought and the influences upon it. His reading appears wildly eclectic, ranging from second-order cybernetics, constructivism, analytical and phenomenological philosophy to Sufism and Buddhism. The contribution of his thought has an outstanding outreach: biology, immunology, autonomous systems, cognitive science, first-person methodologies, and consciousness studies, amongst others. This might appear as a hodgepodge of interests that captured his great mind. However, a deeper look reveals a common theme which expressed itself in a multiplicity of manifestations. The main question that preoccupied him throughout was *What is Life*?

In his search for the answer, Varela moved from the evident to the hidden, making the hidden evident in the process. His starting point was theoretical reflection on the natural systems, and so *autopoiesis* was born (Varela, Maturana, and Uribe, 1974, Biosystems, 5(4), 187-196). *Autopoiesis* [from Greek αὐτο- (auto-), meaning 'self', and ποίησις (poiesis), meaning 'creation'] does not simply mean a system that reproduces and maintains itself in its physical form (boundary). *Autopoiesis* entails emergence of the self and, with it, the world through the process of 'sense-making' driven by the re-actions of attraction and repulsion to the aspects of the environment that have valence (meaning) in terms of organism's function (Thompson, 2004, Phenomenology and Cognitive Sciences, 3(4), 381-398). The perceptual and behavioural (sensorimotor) 'sense-making' habits are shaped and constrained by the biological/neural processes, which are in turn reinforced by the 'sense-making' habits. Like the Ouroboros that continually consumes itself, a living system continuously creates and re-creates its 'reality', for better or worse. In the language of Buddhism, autopoiesis is a biological description of the creation of the species-specific and individual karmic patterns that structure one's view of 'reality', with the strongest habit of all: a perceptual structure of self and other, subject and object, which in the case of humans becomes a pervasive cognitive habit. The perceptual habit which serves a useful functional purpose of biological survival becomes reified into 'I'/'me' and the 'world', at which point we are not just *enacting* our world, but we are 'eveiling' it (notice what happens if the second 'e' is dropped). According to Buddhism, this process of reification (grasping) or fixation on the aspects of our experience is the root of all 'evil' (ignorance of how things really are).

We do not know what happens for a simple organism, like a bacterium or a single cell, but we know for us humans this strong perceptual habit and reification of subject and object results in naïve realism: there is a world out there with the objects that obey the laws of physics and retain their properties whether we are there to perceive them or not, and that our senses provide us with perception of the objects as they really are. With some education, we can revise it to a representationalist view, where we understand that our experience of objects is our internal representation of them. But, in this view, there are still objects out there 'hitting' our senses in different ways according to the laws of physics.

The shortcomings of representationalism are tackled in detail in *The Tree of Knowledge: The Biological Roots of Human Understanding* (Maturana and Varela, 1987, Shambala Publications). In short, the so called 'objects' experienced by the 'subject' are not representations of some features of the environment, but the patterns of neural firing - the perturbations of the organism itself. On the surface of things, this sounds like a materialist statement reducing subjective experience to neural processes. However, Varela in his later work turns the tables completely with the aid of continental phenomenology and Buddhist philosophy of the Middle Way (Madhyamika).

The hidden is brought into the evident in one of his most influential books *The Embodied Mind* (TEM) co-authored with Evan Thompson and Eleonor Rosch (1991, MIT Press). A paradigm of *enaction* formulated in TEM as an alternative to representationalism was well received and is being gradually assimilated by the field of AI and robotics. However, its more subtle but most potent message often gets missed or misinterpreted in the post-TEM formulations of 'enactivist' approach, rendering it flat and innocuous from shaking us out of naïve realism, which continues to haunt us perceptually throughout life and shapes our science as an expression of this dualistic perception despite conceptual education to the contrary. (For an articulation of the differences between *enaction* and 'enactivism' see Vörös and Bitbol, 2017, Constructivist Foundations, 13(1), 31-40).

This subtle and potent message gets reiterated again in a new form in Varela's paper *Neurophenomenology: A methodological remedy for the hard problem* (1996, Journal of Consciousness Studies, 3(4), 330-349). The 'hard problem' as formulated by Chalmers is how and why subjective experience arises from the brain (body) processes. Chalmer's argues that experience is "not an

explanatory posit, but an explanandum in its own right, and so it is not a candidate for [reductive] elimination". Some 'extra ingredient' is necessary to account for how (and why) experience arises from the workings of the brain. What is needed is a non-reductive explanation. Varela agrees, but he proposes a "remedy", rather than an ontological 'solution'. His special ingredient is not something 'extra' that Chalmers calls for.

This special ingredient is not just *methodological*. Varela's statement: "Instead of finding 'extra ingredients' to account for how consciousness emerges from matter and brain, my proposal reframes the question to that of finding meaningful bridges between *two irreducible phenomenal domains*." [italics mine] is often interpreted as simply a pragmatic and paradigmatic proposal of using first-person data alongside third-person (e.g. neuroimaging) data in cognitive neuroscience. The key, however, is that the third-person data, whether brain's neural dynamic or body's physiological process, is framed here as another *phenomenal domain*. In other words, the true problem with the 'hard problem' is a lack of recognition that the brains and bodies are yet another aspect of our phenomenal experience, in the same way that our thoughts, emotions and the sense of self are. We will do well to use our methodological tools afforded by technological progress for the empirical study of the relationships between different aspects of our experience, but we err when we reify (objectify) the 'objects' of our scientific investigation. A methodological reduction as an empirical strategy is necessary and productive. In Varela's words (1976, Coevolution Quarterly, p. 64): "…for every system there is an environment which can (if we so decide) be looked at as a larger whole where the initial system participates. Since it would be impractical to do this at all times, we often chop out our system of interest, and put all the rest in the background as "environment"… To do this on purpose is quite useful; to forget that we did so is quite dangerous".

However, what is even more dangerous, and is the root of the hard problem, is an ontological reduction, either materialist or idealist, or any other form of ontological monism or dualism for that matter, born out of reification of a particular aspect of our experience that we have become fixated upon for the purpose and in the process of analytical or empirical investigation. Varela's remedy for the hard problem echoes Wittgenstein's "treatment" of "philosophical illness" – a shift in perspective that reveals the incorrectness of assumptions giving rise to a perceived "problem" (Bitbol, 2012, Constructivist Foundations, 7(3), 165-173). Seeing through our presuppositions born out of reification dissolves (rather than solves) the hard problem and reveals what is hidden by always being in plain view – the primacy of experience or awareness. "Lived experience is where we start from and where we all must link back to, like a guiding thread" (Varela 1996, Journal of Consciousness Studies, 3(4), p. 334).

This radicality of Neurophenomenology is often overlooked (Bitbol and Antonova, 2016, Constructivist Foundations, 11(2), 354-356). It is not enough to understand the point conceptually. The true potency of the remedy is only achieved through the personal 'mutation'. To swallow Varela's pill sincerely and wholeheartedly requires deep personal commitment to the transformation of one's conscious experience and its application to all life, including our scientific pursuits.

Why a commitment? Because of our deeply engrained habit of reification that required repeated 'treatment', over and over again, every time it surfaces. Why is it difficult to apply? Because of our tendency for "Cartesian Anxiety" – our need for a fixed foundation for knowledge, and absolute stable ground, without which we fall into chaos due to our low tolerance of uncertainty. We have low capacity for sitting with the unknown. Instead of staying with the question, we fall for an answer, however problematic. The middle way between absolutism and nihilism is the stance of "groundlessness" (Varela, Thompson, and Rosch, 1991). In TEM, "groundlessness" is argued from the Madhyamika's point of view. It is not only a conceptual doctrine, but an experiential stance, *a way of being*, free of reification of any kind.

The growing field of Contemplative Neuroscience is starting to provide us with glimpses as to what happens to the neural dynamics of meditation practitioners who are able to rest in "groundlessness" for periods of time. Experienced mindfulness practitioners show a weaker formation of perceptual habits as demonstrated by reduced habituation to repeated auditory startles (Antonova et al., 2015, PLoS One, 10(5):e0123512); less predictability of neural dynamics (Nehaniv and Antonova, 2017, IEEE Symposium Series on Computational Intelligence, 1753-1761); greater speed and fluidity of neural dynamics during semantic processing (Pagnoni et al., 2008, PloS One, 3(9):e3083), and decoupling of sensory experience of pain from its affective component (Grant et al., 2011, Pain, 152(1), 150-156) supporting the dictum: "Pain is inevitable. Suffering is optional." (Haruki Murakami). Relaxing the reification of the 'self' as 'I' or 'me' in the state of mindfulness attenuates the activity of the Default Mode Network (DMN) (for more details on 'self' and the DMN with the implication for AI/robotics see Antonova and Nehaniv in this volume).

In conclusion, I would like to propose a new circularity – that between the fields of AI/robotics and Contemplative Neuroscience (CN). CN could and should become a new source for AI/robotics in informing on cognitive and neural dynamics of *enaction* that do not lead to 'eveiling'. AI/robotics and computational neuroscience could become a testing ground for neural (or other) models of "groundless" cognition, assisting humans in achieving de-reification through brain-computer interfaces or immersive experiences such as created in a fascinating exhibition *Artist and Robots* in Paris exploring Artificial Imagination of AI (https://www.grandpalais.fr/en/event/artists-robots).

And finally, my challenge as a contemplative neuroscientist to my AI/robotics colleagues: can we build an Artificial Buddha that functions out of wisdom and compassion for the benefit of all beings? Who experiences and cognizes *Life* as an eternal dance of polarities rather than a perpetual war of dualistic opposites brought about by reification of basic structures of experience; *ALife being* that at all times follows three basic Buddhist instructions for *Life* free of suffering: non-grasping, non-grasping, non-grasping.