

# Neurophenomenology, an Ongoing Practice of/in Consciousness

Michel Bitbol • CNRS, France • [michel.bitbol@polytechnique.edu](mailto:michel.bitbol@polytechnique.edu)

**> Context** • In his work on neurophenomenology, the late Francisco Varela overtly tackled the well-known “hard problem” of the (physical) origin of phenomenal consciousness. **> Problem** • Did he have a theory for solving this problem? No, he declared, only a “remedy.” Yet this declaration has been overlooked: Varela has been considered (successively or simultaneously) as an idealist, a dualist, or an identity theorist. **> Results** • These primarily theoretical characterizations of Varela’s position are first shown to be incorrect. Then it is argued that there exists a stance (let’s call it the Varelian stance) in which the problem of the physical origin of primary consciousness, or pure experience, does not even arise. **> Implications** • The nature of the “hard problem” of consciousness is changed from an intellectual puzzle to an existential option. **> Constructivist content** • The role of ontological prejudice about what the world is made of (a prejudice that determines the very form of the “hard problem” as the issue of the origin of consciousness out of a pre-existing material organization) is downplayed, and methodologies and attitudes are put to the fore. **> Key words** • First-person approaches, non-dualism, idealism, Francisco Varela, Edmund Husserl.

## Introduction

When he elaborated the discipline he called “neurophenomenology,” Francisco Varela was explicitly looking for a way out of the well-known “hard problem” of the physical origin of consciousness (Chalmers 1996, 1997). Overlooking the subtitle of his original paper (Varela 1996), which prudently announces a “methodological remedy,” one may figure out that he was seeking a proper *solution* to this problem. Along with this construal of the scope of neurophenomenology, it is natural to attempt a characterization of it as a *theory* of consciousness, and to pick out similarities between this alleged theory and other classical theories. Wandering then in the writings of Francisco Varela, one may find kinship between his views and either idealism, dualism, or mind-brain identity theory (in succession or even simultaneously). But whenever one of these comparisons is pushed too far, it tends to collapse (Bitbol 2000, 2002, 2006). At any rate, the very multiplicity of such comparisons strongly suggests that none of them taken in isolation can claim to capture the essence of Varela’s strategy when dealing with the hard problem of consciousness. We are then left with a sort of enigma: What was, after all, Varela’s position on the status of consciousness in the world? Was there any such position at all?

The hypothesis that will be formulated at the end of this article is that there is a perfectly consistent Varelian position on consciousness, but that this position has nothing to do with a *theory*. Any theory is successively examined and rejected, in the style of Wittgenstein and of *Madhyamaka* (the Buddhist *Middle Way*), and what is left is a pure *stance*, which has both a lived and a professional aspect. In this stance, the problem of the origin of consciousness becomes irrelevant; it is rather absorbed into a way of living and a program of research. No “solution” is given to this problem, but a mere “dissolution” that has both an existential and a methodological component.

Accordingly, I will first examine (and reject) three possible doctrinal characterizations of Varela’s position on consciousness. In the final section, I will state what is the proper stance in which the problem of the physical origin of primary consciousness (or pure experience), as contrasted to reflective consciousness, does not even have to arise.

## Idealism?

A recurrent theme over thirty years of Varela’s writing is a statement of epistemological and sometimes ontological primacy of mind that can easily be mistaken for full-

blown idealism. Varela’s shortest formulation of the Archimedean point of any study of consciousness can be found in an early article. According to him, *by “Being” one must understand nothing else than experience* (Varela 1976). Being is experience, experience is Being. Conscious experience here does not reduce to a mere appearance that can be overcome to reach a thinkable reality. Rather, experience is construed as *the unique and insuperable reality*; the reality out of which any intellectual production arises, including the intellectual production that supports our theories of neural functioning.

Much later, in the second half of the 1990s, Varela continues to build on this ground to elaborate his neurophenomenology. He first points out that “the phenomenological approach starts from the irreducible nature of conscious experience.” And he endorses this starting point by claiming that “Lived experience is where we start from and where we all must link back to, like a guiding thread” (Varela 1996, 1999a). This deliberate decision in favor of an experiential starting point is clearly in contrast to the leading presupposition of analytic philosophers who formulated the standard version of the problem of consciousness. Their presupposition is that one must take an objective stance from the beginning, then try to show how subjective experience

may emerge from a set of physical objects considered as preexistent and real. The neurophenomenologist then tends to accuse this very standard (analytic) formulation of being the true reason for the existence of a problem and of its stubbornness. According to him, the path towards a progressive resolution of difficulties is bound to be run through backwards.

What should we think of this departure point (Taraborelli & Mossio 2008)? Has it not already been evaluated and rejected in the history of thought? After all, insisting on the priority (and even sometimes on the *ontological* primacy) of the experiential domain over the physical domain is by no means new. William James thus advocated the absolute antecedence of “Pure experience” in his *Essays of Radical Empiricism* (James 2003). He declared that the very actuality of the so-called “external” objects is numerically one with part of our lived experience, and that “pure experience” is then to be taken as the “neutral domain” out of which everything else arises. However, James was soon accused of having merely revived old-fashioned idealism; a form of idealism he had borrowed from Berkeley, whom he often quoted, and from the post-Kantian philosophers with which he was familiar. More recently, other varieties of this position have been formulated and presented as alternatives to contemporary physicalism. One of these varieties was expressed by Douglas Bilodeau, who concludes a long study of the relations between the problem of consciousness and the philosophy of modern physics by stating his “intuition” according to which *consciousness is much closer to the fundamental level of reality than anything meant by physical concepts* (Bilodeau 1996). A systematic attempt in this direction is Piet Hut’s, who proposes to overturn the “hard problem” of consciousness by starting the inquiry from experience rather than from its hypothetical physical substrates (Hut & Shepard 1998). But here again, the reaction of the community of researchers (be they scientists or philosophers) looks like a reproachful silence. The idea of reverting the order of methodological or ontological priorities gained ground only recently in a prudent phenomenological formulation (Thompson 2007; Gallagher & Zahavi 2008).

After all, this usually indifferent or even hostile attitude is understandable. What can we do next when we have come back to the experiential source of any assertion, including physicalist propositions? Can we move forward from this point, or are we doomed to stagnation, contenting ourselves with repeatedly chanting that experience is Being, that experience is the ultimate source, that pure experience is everything? Is the physicalist departure point not more fruitful, not of course to solve the “hard problem,” which is likely to remain stuck in the blind spot of this doctrine, but at least to solve so many other accessible problems whose scientific interest is obvious? And was Carnap not convinced of this as early as 1928, when he progressively abandoned the project of basing his *The Logical Structure of the World* on an “auto-psychological basis” and on “methodological solipsism,” considering that it was better to come back to a more common physicalist basis?

These two misunderstandings must be overcome if we want to realize that Varela’s insistence on the primacy of experience is no mere expression of an old-fashioned position, but rather a sign of his being far ahead of the present stage of scientific and philosophical research. Firstly, his position is not to be mixed up with a full-blown version of idealism. And secondly, far from overlooking the issue of fruitfulness, he brings his work to an exceptional level of efficiency.

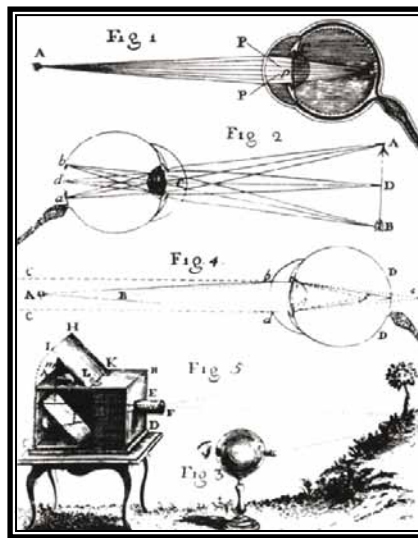
Let’s start with the first misunderstanding. Pulling back every problem to its status in lived experience is definitely *not* tantamount to an idealist doctrine. The reason why this looks so difficult to understand is the following: when one tries to characterize the position held by Varela in the field of the philosophy of mind and consciousness, the said characterization is automatically underpinned by the very dualistic prejudice he is trying to overcome. Accordingly, one feels that, between the physical and the psychical, Varela has chosen the psychical; and that the psychical is taken by him as some sort of ontological ground. But this is twice wrong. He did *not* choose the psychical *qua* opposed to the physical; and he soon avoided the strong ontological commitment of his early writings in favor of a purely operational standpoint.

Indeed, the starting point of Varela’s neurophenomenology is no abstract inter-nal realm, but lived experience in its entirety, human life in its full depth and extent. The starting point of neurophenomenology is *embodied human life*, embedded in an own-body which is both seeing and seen, and is thereby inextricably connected to an environment made of alter-egos and inert objects. This integral human life has in it all the indispensable resources to objectify fragments of what occurs; and conversely, it has the possibility to draw a circle of interiority by retaining in it the residue of objectification. But human lived experience can by no means be reduced to this circle of interiority, which is one of the end-products of its own discriminative work. Lived embodied experience does not reduce to some “inner”/“psychical” domain, insofar as it cannot be contrasted with a physical domain that still has to be constituted in it and by it. Lived experience is, in fact, the common ground from which every separation between subject and object arises. In lived experience, embodiment is no mere additional feature. Embodiment is the central axis that unites experience and its objective correlates, and which gives simultaneous access to the realm of phenomenology and to the experimental material of the cognitive sciences (Varela 1997; Rudrauf et al. 2003; Lutz & Thompson 2003). Lived experience – embodied, empathic, interactive, and even verbalized experience – is the mixed and still indistinct ground out of which one can constitute progressively both the domain of pure subjective phenomena and the science of nature (Merleau-Ponty 1964).

Varela is even very careful in his project of neurophenomenology to avoid the overtones of idealistic *ontology* that could still be heard in his first writings. His leit-motiv in neurophenomenology has become universally anti-foundationalist. There is no foundation in the objectified entities of physics; but no ultimate foundation either in the field of subjectivity: “Exploring first-person accounts is not the same as claiming that first-person accounts have some kind of privileged access to experience. No presumption of anything incorrigible, final, easy or apodictic about subjective phenomena needs to be made here, and to assume otherwise is to confuse the immediate character of

the givenness of subjective phenomena with their mode of *constitution* and *evaluation*" (Varela & Shear 1999: 2). Further on, Varela makes it clear that "No methodological approach to experience is neutral; it inevitably introduces an interpretative framework into its gathering of phenomenal data. To the extent that this is so, the hermeneutical dimension of the process is inescapable: every examination is an interpretation" (Varela & Shear 1999: 14). An interpretative framework is already imposed on experience by the language that serves to express it in first-person reports. Moreover, writes Varela, thus bringing us to the heart of the neurophenomenological project, a second interpretative layer can be added by comparing these first-person expressions with the objective data of neurophysiology. In neurophenomenology, neurophysiology can indeed be used as an interpretative filter of first-person experience just as much as, conversely, first-person reports are crucial to interpreting raw neurophysiological data. Each one of these interpretative layers must not be regretted as a loss of immediacy, but rather greeted as an advance in the procedure of constitution by which a new compound science of conscious experience is elaborated.

This strategy yields what we may call a balanced science of consciousness. Neither first-person reports nor objective structures are considered as fundamental. Neither of them is more fundamental than the other. Instead, each side is suspended in an interpretative scheme or a process of meaning-ascription drawn from the other side. The balanced strategy implies no physicalist foundationalism, according to which conscious experience is nothing else than an emergent property, or an epiphenomenon, of some object of the science of nature (Bitbol 2007). Neither does it imply an idealist foundationalism, according to which subjective experience is an ultimate *given* out of which everything else, including objective nature, is elaborated. True, the process of *constitution of objectivity* that has repeatedly been evoked until now may sound like foundationalist or dogmatic idealism. But this is only a superficial judgment. Indeed, constituting objects in the sense of Kant and Husserl does *not* amount to projecting matter out of subjective stuff (which would indeed be tantamount to dogmatic idealism), but



**Figure 1:** Left: The eye from a third-person (naturalized) standpoint. Right: The eye from the first-person (transcendental) standpoint.

rather delineates a process of *organization* of lived experience by way of concepts (e.g., Natorp 1912). So, instead of either physicalist or idealist foundationalisms, what is assumed by neurophenomenology is a permanent circulation from one side to the other side, a feedback loop uniting the first-person and the third-person approaches within a human, embodied, situated, social form of life shared by researchers.

This brings us closer to the answer Varela gave to the second misunderstanding: the claim that he contents himself with declaring that conscious experience is irreducible, and that he thus loses the fruitfulness that characterizes the cognitive sciences and the science of nature in general. Actually, Varela was perfectly aware of this problem, and he reproached other authors for ignoring it. He started with quoting approvingly a sentence of John Searle, according to whom *one cannot represent subjectivity as part of our representation of the world because subjectivity is nothing else than this representation itself* (Searle 1992: 98). This wonderful oracular remark echoes another one by Erwin Schrödinger (1967: 138): "The reason why our sentient, percipient and thinking ego is met nowhere within our scientific world picture can easily be indicated in seven words: because it is itself that world picture" (Figure 1). But immediately after this, Varela

noticed that as soon as Searle has successfully advocated the philosophical thesis of the irreducibility of conscious experience, he is unable to draw any conclusion from it. In particular, he offers no clue as to how to solve the epistemological puzzle of the study of a self-adherent non-object such as consciousness (Varela 1997). Searle was then accused by Varela of practicing pure negative epistemology, thus being doomed to sterility. By contrast, Varela never lost his interest in the fruitfulness of the science of nature. He personally contributed to the advance of neurophysiology with his pioneering work on long-distance correlations in neural assemblies of the brain cortex. And he considerably reinforced its reliability by complementing it with a dynamical feedback between first-person reports and third-person descriptions. Important discoveries concerning the anticipation of epileptic seizures by coupling experiential self-monitoring and neural pattern detections were made this way (Petitmengin, Navarro & Baulac 2006). In fact, Varela demonstrated that it is by ignoring first-person experience that one imposes limits on the fruitfulness of neuroscience, while these limits can easily be overcome as soon as a circulation between the first and the third person is established. But how does this feedback work, and why does it provide additional efficiency?

To begin with, we must note that nobody, not even a hard-nosed eliminativist, has completely ignored this sort of feedback. Studies of neurophysiological experiments can hardly be taken as inquiries about the *mind* if they have never been compared with something like a first-person report. A cortical area, for instance, is characterized as “visual” by virtue of a correlation between its activation and a report of visual experience. True, when purely sensori-motor activities are involved, the report can be avoided in certain circumstances and replaced with an evaluation of behavior in response to colors, movements, or shapes. But as soon as higher-order mental activities, such as imagination or thought, intervene, this possibility of replacement vanishes. True, also, where animal neurophysiology is concerned, the report is simplified or even skipped and replaced with behavior. But here again, it makes sense to declare that the exploration of animal *mind* is undertaken only by due analogy with lived and reported mental processes. Moreover, in every case, the categorization and fine-grained resolution of cortical activities is greatly improved by refinement of reports. As Jean-Philippe Lachaux (2011) cogently points out, taking first-person reports carefully into account is our only way to extract information from a host of cortical events that would otherwise be categorized as “neural noise.”

In spite of all this, as soon as a few first-person/third-person correlations have been brought out, one just takes them as granted and then marginalizes them in favor of a virtually exclusive study of neurophysiology. Indeed, it is widely believed that neurophysiology is the only proper field for making decisive *advances* in the science of mind. From a quasi-exclusive focus on neurophysiology to eliminativism, there is only one step: the short step that separates research strategies from ontological commitment.

Varela fully takes into account this procedural feature, which is both neglected and universal. But, unlike those researchers who use it unreflectively, he purports to push the first/third person mutual interaction to its maximal level of efficiency and fruitfulness. To reach this aim, he writes, one must reinforce the mutual shaping between the two regions of knowledge. One must not

be restricted to an inventory of concomitances that remain quite approximate as long as the quality, stability, and reliability of subjective reports remains much lower than that of neurophysiological data. One must rather replace the simple statement of a concomitance of events with “mutual constraints” between two lines of inquiry. And, moreover, these two lines of inquiry must both be pushed at the same degree of exactingness (Varela 1996). The connection between the first and the third person approach is then no longer static but dynamic, no longer logical but methodological (Varela 1997). Once this is done, the connection between methodology and ontology produces radically new results. In so far as, from now on, the first-person approach is taken as just as important – and just as liable to improvement (Petitmengin & Bitbol 2009) – as the objective study of neural processes, no absolute primacy of the physical over the mental has to be postulated. The twofold focus of research precludes the unique focus of ontology. “*Not one*,” to paraphrase the first part of the title of Varela’s early paper (Varela 1976).

## Dualism

Should we then say, positively, that the twofold focus of methodology imposes a twofold focus of ontology? Should we suspect that Varela’s position in fact reduces to a variety of dualism (Bayne 2004)? This does not follow: “*Not two*,” to paraphrase the second part of the title of the same paper. But in order to clarify this point, we must inquire upstream, closer to the historical root of dualism, namely to René Descartes.

To begin with, it is very important to note that Descartes’ dualism was not born as an ontology (the two-terms ontology of *res cogitans* and *res extensa*). Dualism should then not immediately be taken as a “mistake,” unlike what Damasio (1995) and so many other authors have claimed. In fact, as the first two *Meditations* of Descartes show, dualism arose from a radical inquiry into lived experience (Henry 1985) and into its ability to ground the kind of truth required by a science of nature. It thus appears that dualism was the best possible initial compromise between the need to es-

tablish a sphere of objectivity for the birth of the science of nature, and the need to keep hold of lived experience (or pure appearance) as the ultimate background of every claim of knowledge and every program of systematic objectification. Of course, this initial compromise was eagerly criticized. But most attempts at discarding the terms of the Cartesian compromise yielded a lopsided position, a distortion on one side or the other. La Mettrie’s or Diderot’s materialism tended to glorify science, yet become blind to its very source, which lies in what Husserl called the “life world.” Conversely, the reaction of romanticism stirred up passion for the richness of experience, for its resonances with nature and its potential for art, yet it tended to discard the major methodological options that were crucial for modern science (Elie 1993).

The Cartesian arrangement was then remarkably useful in order to build a conquering science that still retains contact with its experiential humus; a science assured of itself by its own achievements yet not forgetful of its own condition of possibility. Simultaneously, however, the substantialist features of dualism (what we may call its “ontological crystallization”) prepared the future excesses that consist of granting exclusivity to one of its poles and remaining stuck in one of its alternative stances. So, if one wishes to overcome the reifying tendency of Cartesian dualism without losing its balanced approach, it is crucial to acknowledge its initial motivation and to find a good substitute for it.

The right strategy at this point is to start once again from Descartes’ ability to adopt successively two stances with regard to experience, namely a reflective phenomenological stance on the one hand, and an intentional objectifying stance on the other hand. Indeed, the feature of dualism that must be overcome is not the duality of stances but its substantialist transcription, not the oscillation between two postures but its recurrent paralysis.

Neurophenomenology is an appropriate response to this challenge. For it represents a dual yet not dualist approach to the problem of mind, consciousness, and nature. It renews the genuine motivation of Cartesian thought without falling into its metaphysical pitfalls.



Varela identified very early the method to be adopted to achieve this aim. According to him, a *mutation of experience* (of *Being-there*) is just as important as an intellectual advance if one wishes to overcome the mind-body dualism (Varela 1976). In other words, the mind-body problem (and the problem of consciousness as well) is not the kind of problem that can be separated from us and treated as an abstract object of our thought, but rather a problem in which we are so inextricably involved that it can be addressed only at the cost of transforming ourselves. This remark is very much in tune with Descartes and with the phenomenological lineage that explicitly followed him (Husserl 1977), for it advocates a systematic work *on* and *in* experience, not (or not only) *about* experience. However, the sought mutation of experience is not restricted to an ability to rehearse the act of phenomenological reduction. It rather consists in favoring the plasticity and interchangeability of the modes of attention and of the associated directions of investigation. This institutes a productive spiral of neurophysiological and phenomenological categories. Here, no attempt is made to circumscribe the pre-set terms of a dualist scheme. Rather, a strategy of joint advances, comparative heuristics and mutual assistance between two basic stances is instituted.

Accordingly, the “hard problem” of the physical origin of consciousness is not solved but rather put to rest, or dispelled (Bitbol 2000, 2002, 2008a, 2008b; Peschard & Bitbol 2008). From a neurophenomenological standpoint, the issue of the physical and neurophysiological origin of consciousness is a non-starter, a question that does not even have to be formulated. Indeed, formulating this question presupposes that one has endowed the domain of physics with a status of fundamental Being, whereas this domain only represents one step in an ongoing dialectic of subjectivation and objectivation, of lived experience and its stabilized contents.

This seems epistemologically simple, obvious; almost *too* obvious. So much so that Varela’s dissolution of the problem of consciousness remains misunderstood. Many specialists in cognitive science and philosophers of mind (e.g., John Perry) have said to me in private discussions that they consider it as a mere dodge.

This reluctance is another point of similarity between Varela’s neurophenomenology and Descartes’ dualism: few philosophers, and even fewer cognitive scientists, realize that the duality of *res cogitans* and *res extensa* is only a late solidification of an ongoing dialectic of (i) reflectivity about consciousness and (ii) doubt about intentional objects that (as stated in the two first *Metaphysical Meditations*) entirely develops within lived experience. These philosophers therefore demand a proper explanation of the communication between the two substances, and they consider that Descartes is dodging their requirement when he insists, in a letter to Princess Elisabeth, that “everyone invariably *experiences* (the union of mind and body) in himself *without philosophizing*” (Descartes 1991: 228; Bitbol-Hespériès 2000). When he writes this sentence, Descartes is by no means cheating, since he is faithful to his phenomenological starting point. He is not forgetful of the fact that the connection between embodied experience on the one hand, and the body as object of experience on the other hand, is itself an *experience*. Descartes is therefore more coherent than those who require from him something like a *physical* (!) explanation of the connection between the physical and the mental realms.

Returning to Varela, the reason for the misunderstanding about his approach to the hard problem is likely to be that his attempt is truly revolutionary, just as Descartes’ *tabula rasa* was in his time. Firstly, Varela sketches a radically new conception of science. He extends science not only in its ontology, by adding new objects or new properties (like Chalmers), but in its very methodological definition. Indeed, from now on the method of science encompasses the rules of exchange and communication of the subject’s standpoints. It is no longer restricted to the objective byproduct of these rules, which is an invariant relative to any change in the subject’s standpoint. Secondly, the permanent interplay between first-person and third-person descriptions that is taken by Varela as the very root of his approach is hardly accessible from ordinary states of consciousness. The so-called “natural (or intentional) attitude,” which is the dominant state of consciousness in our civilization, imposes a biased perception of

the neurophenomenological feedback loop. It dictates an unbalanced view according to which a first-person description can only be a sort of deteriorated expression of what *really* occurs in the brain cortex. It is then only in an amplified, flexible, expanded state of consciousness, such as that which was systematically cultivated by Varela in his practice of meditation and phenomenological reduction, that the two stances (reflective and intentional) become truly symbiotic and offer a balanced perception of the first and third person approaches. It is only in such a state of consciousness that no temptation may arise to *reify* the byproduct of the two stances in order to fit with the intentional norm of knowledge of our civilization, which can only commit itself to objects or things (*res*).

## Identity theory?

Yet the list of misunderstandings is not complete at this point. Varela was so at ease with the interplay of stances and approaches, and so serious in the third-person investigations he performed as a biologist, that he was sometimes accused of being a hidden *identity theorist*.

By identity theory, what is usually meant is a highly biased view in which mental processes are allegedly identical to neurological processes, and should not therefore be distinguished from them (Place 1956; Armstrong 1993). This is tantamount to a one-way identification in which the mental is identified with the neural, but not (at least not explicitly) the other way around. But actually, in one of the major original versions of the identity theory defended by Herbert Feigl, things were more intricate and more interesting. In this original version, the scheme of identification is strictly symmetrical: it works in both directions. The physical is nothing else than the mental, just as much as the mental is nothing else than the physical. This balanced view is underpinned by a two-aspect monist metaphysics: “Instead of conceiving of two realms or two concomitant types of events, we have only one reality which is represented in two different conceptual systems – on the one hand, that of physics and on the other hand, where applicable, that of phenomenological

psychology" (Feigl 1970). The only privilege granted to physics by the identity theorist is not to reveal the ultimate essence of reality, but only to pick out regularities within it. After all, "this reality is known to us by acquaintance only in the case of our direct experience" (Feigl 1970).

So, why was Varela so often mistaken by some philosophers for an identity theorist (Lockwood 1993)? The reason for this conflation is quite simple to understand. Firstly, as we have just seen, both Varela and the identity theorists resist any form of *ontological dualism*, including Chalmers' property dualism, and they discard accordingly the picture of the so-called "Cartesian theater" (which has in fact little to do with Descartes). Secondly, both Varela and the identity theorists consider that the relations between mental and neural processes are stronger and more reciprocal than in Davidson's anomalous monism (Thompson & Varela 2001; Davidson 1980). Moreover, from a semantic standpoint, Varela and his collaborators were not far from considering that, in the future, fixing the meaning of certain delicate and discriminating phenomenological "descriptions" could depend in a crucial way on their disciplined correlation with neural events. Thus, in his long and careful comparative study between the phenomenology of time-consciousness and the time-scales of neurological processes, Varela often uses the language of neural dynamics to buttress the phenomenological descriptions drawn from Husserl. He especially examines the concept of "double intentionality" proposed by Husserl (1964). According to Husserl, in the experience of time there is both tension towards the past and the future and tension towards the self-manifestation of this polarized structure. Varela then asserts that this complex intentional structure "... is of this class of dynamical bootstrap" between "the phase space landscape and the specific trajectory that moves in it" (Varela 1999b). The "is" of identification is used here, not the more prudent "is correlated with." An experience as described by the phenomenologist is apparently *identified with* a dynamical structure of neural functioning. The feeling that Varela advocates a sort of theory of identity, and even a biased form of it, is reinforced further when he invokes elementary neural events as the "*biological*

*basis*" of the experience of time, or when he insists on *naturalizing* the phenomenology of time consciousness with its typical retention/protection structure. Insofar as this combined study of the phenomenology of time-experience and neural dynamics was taken by him as the "acid test" of the validity of a neurophenomenological program of research, should we not conclude that what he was actually purporting to show was the validity of a mind-brain identity theory?

Some authors (Rudrauf et al. 2003) even suggested that, along with his studies of brain dynamics, Varela had sketched a true neurological or biophysical theory of the arising of consciousness, thus attempting to overcome the "explanatory gap." Had he not declared that affectivity, emotion, is "*generative for consciousness itself*" (Varela & Depraz 2000)?

As a consequence, Varela was well aware that one might ask him: "Is this not just a fleshed-up version of the well-known identity theory?" (Varela 1998). But he answered the question by pointing out that in his approach, theoretical matters are systematically deflected onto a methodological plane. His neurophenomenology is not an identity *theory* of some factually given neuro-experiential correlation; it is a *procedure* of systematic *institution* of such relationship, and of correlative refinement of the phenomenological terminology. Varela here implicitly expanded Wittgenstein's "grammatical" analysis of expression. Wittgenstein restricted his investigation to the way the standard norm of interconvertibility between (first-person) expressions and (third-person) descriptions of external behavior institutes an intersubjectively acceptable folk-psychological vocabulary. And Varela amplified his field of interest to a norm of mutual constraint between (first- or second-person) phenomenological descriptions of stabilized contents of experience and (third-person) neuroscientific reports. While in Wittgenstein's work, the form of life in which the use of expressive sentences makes sense reduces to our everyday activity, in Varela's work, the relevant form of life is broadened so as to include disciplined practice of phenomenological reduction and neuroscientific experimenting and/or theorizing as well. For instance, what Varela was after in the article in which he claimed

to have disclosed the emotive root of consciousness was not a theory of the absolute origin of pure experience or "sentience," but only a neurological counterpart of the "specious present," namely of a short-termed stabilized, unified, and highly reflective type of consciousness act.

We now see that in no way can Varela's position be assimilated to a blend of identity theory. Rather, identity theory can be characterized as one more reifying and dissymmetric account of the ongoing dialectic of embodied experience and objective reports that Wittgenstein displayed in ordinary life and that Varela extrapolated to a refined combination of experiential and scientific forms of life. This theoretical account is reifying because, in its most biased form, it usually takes for granted that objective reports (of neural events) disclose *things as they are*. And it is dissymmetric because, even though it relies more or less tacitly on a background of first-person experience, this background is overlooked in favor of a claim of ontological primacy of third-person descriptions of neural events. The kind of "solution" to the mind-body problem offered by identity theories is tantamount to a curtailed and unbalanced variety of Varela's dissolution.

It is quite easy to detect in Varela's writings a confirmation that the dialectical method of neurophenomenology is endowed with more value than any theory that could be derived from it. In his article about the physiology and phenomenology of time-consciousness, which displayed so many signs of apparent sympathy for the theory of identity, the true outcome of the investigation is by no means couched exclusively in the language of neurology. Rather, this outcome is a general scheme called "the four-fold structure of *nowness*" that inextricably expresses the information drawn from phenomenology and neural dynamics. In fact, the latter scheme is neither phenomenological nor dynamical, but derived from the mutual constraints exerted by both domains on one another. It represents a sort of hybrid structure that can be taken alternatively as phenomenological and dynamical, by virtue of a set of translation rules established during the feedback process that is typical of Varela's neurophenomenological method. True, the fact that there *are* translation rules

could be interpreted as an argument in favor of identity theories. But in the framework of neurophenomenology, the opposite is the case. In this framework, the irresistible tendency of some philosophers to slip towards identity theory is to be understood as a solidified expression of the fruitfulness of such translation rules. Once again, method comes first, and ontology is the shadow cast by methodology.

## Conclusion

We have just seen that neurophenomenology is neither an idealism, nor a dualism, nor an "identitism." Francisco Varela's neurophenomenology is no theory of consciousness, and can therefore afford no theoretical "solution" to the "hard problem" of consciousness. Any such theory is rejected in view of its arising from false dichotomies, and what is retained instead is a methodology of research. Just as Wittgenstein rejected any accusation of being a behaviorist, an idealist, or even a pragmatist (because he was immersed in a practice of behavior, of mental life, and of everyday linguistics and pragmatics, instead of holding some theoretical version of these practices), Varela could easily reject any accusation of holding any one of these "isms" because he rather prescribed immersion in a multidimensional practice of phenomenological examination and scientific inquiry. Actually, the very belief that there may exist a *theory* or a set of *conceptual elements* that should solve the problem of the origin of conscious experience is considered as radically misleading by Varela. In functionalism, as in any other theoretical approach to consciousness, "what is missing is not the coherent nature of the explanation but its alienation from human life. Only putting human life back in will erase that absence; not some 'extra ingredient' or profound 'theoretical fix'" (Varela 1996). Recovering the fullness of lived life rather than remaining trapped in a restricted version of it expressed by a theory is the only way of tackling the problem of consciousness.

According to Varela as I understand him, the greatest possible mistake in the philosophy of consciousness is to seek a distanced view of the ongoing dialectic

of commitment and distanciation that underpins the elaboration of an objective picture out of the multiplicity of our situated standpoints. We are plunged into a situated life, and we make an effort to push certain features of this life to a certain distance (by extracting invariants from them) in order to master these features collectively, independently of individual, spatial, and temporal situations. It is then absurd to try to capture in a theory of the objective world the very starting point of the process that allowed objectivity to exist in the first place. It is absurd, in other terms, to purport to elucidate the subjective by a theory of the objective, as if one could play with words and overcome an in-principle difficulty by means of a trick.

So, the only acceptable strategy is to widen again the field of attention, to cease to be exclusively absorbed in intentional objects, and to consider the whole span of human life, with its alternation of reflective awareness of one's present situation and escape towards a realm of theoretical idealities. This way, one may finally "move beyond" the problem of the origin of consciousness by just ignoring theoretical issues, and remedy the "practical ignorance" of those who are overfascinated by these issues (Varela 1996). Practical ignorance is replaced by practical knowledge as soon as one becomes able to articulate two styles of approaches that are inherent to the very workings of our minds: contact with experience on the one hand, and extraction of stable features of this very same experience for the sake of intersubjective agreement on the other hand. Accordingly, the so-called "hardness" of the hard problem boils down to the hardness of practical training, and to the hardness of changing our very conception of science in order to let it encompass its lived source and performative procedures as well as its objects and achievements. This twofold hardness can easily be softened by: (i) due seriousness in training one's mind to the phenomenological exploration of experience and to the versatility that is indispensable to go from there to objective science; (ii) acceptance of a completely renewed and broadened conception of science that includes such training. It is only this way that one will agree to reframe the age-old problem of the origin of consciousness, and to

dissolve it in an appropriate epistemological and existential stance, rather than "trying to solve it within its original setting" (Varela 1996).

To conclude, I wish to compare Varela's dissolution of the problem of consciousness with two other dissolutions: the antimetaphysical positivistic dissolution and the neuro-materialist dissolution. According to the antimetaphysical positivistic dissolution (Carnap 1936, Ryle 1949), the mind-body problem arises from certain logico-linguistic confusions. One thus witnesses a logical confusion between two categories of terms (organizational and objectual; dispositional and categorical), which include, respectively, the domains of mental and material properties. One also witnesses a linguistic confusion that consists in mistaking intersubjectively learned mental terms for names of true inner subjective *entities*. These remarks about language are sound, but they by no means exhaust the difficulty, and they even contribute to it. Indeed, what these authors fail to mention is that the logico-linguistic confusions they denounce are underpinned by an *experiential* confusion: a tendency to be so absorbed in the intentional stance and its natural objects that even the *background lived experience that adopts stances in the first place* is mistaken for one of its objects and named like any other object. The very exclusive focus of logical positivists on logic and language partakes of this ultimate confusion (for logic and language are themselves objects of thought), and it therefore does not alleviate the problem they wanted to dissolve. As for the neuro-materialist dissolution, it amounts to asserting that it is science itself that defines what counts as an explanation of something, and that an increasingly accurate *correlation* between neurophysiological processes and first person reports should count, in its framework, as an explanation of consciousness by neurophysiology (Hardcastle 1996). Unfortunately, the experiential bias of this neuro-materialist dissolution is just as obvious as that of the antimetaphysical positivistic dissolution. Its proponents are so focused on the object of their study (namely neurophysiological processes) that whenever they are able to display the correlation of this object with some first-person report, they consider it



## MICHEL BITBOL

is presently “Directeur de recherche” at the Centre National de la Recherche Scientifique, in Paris (France). He teaches Philosophy of modern physics and Epistemology. He received, successively, his M.D. in 1980, his Ph.D. in physics in 1985, and his “habilitation” in philosophy in 1997 in Paris. He worked as a research scientist in biophysics from 1978 to 1990. From 1990 onwards, he turned to the philosophy of physics. He edited texts by Erwin Schrödinger and developed a neo-Kantian philosophy of quantum mechanics. He also studied the relations between the philosophy of physics and the philosophy of mind, working in close collaboration with Francisco Varela. He is recipient of an award from the “Académie des sciences morales et politiques.”

as the material *origin* and *explanation* of the conscious experience that is reported. They do not even realize that, correlation being a two-way relation, the decision of taking one of its terms as *explanans* and the other as *explanandum* (rather than the other way around) is arbitrary. As Bas Van Fraassen (1980: 141) pointed out, any explanation is an answer to a why-question, and why-questions are entirely shaped by prejudices about what is taken for granted and what is considered to be in need of explanatory justification. Like any other, *this* decision about the direction of explanation arises from a prejudice, according to which objective neurophysiological processes are to be taken for granted, and what needs an explanatory justification is the very experience that silently performs the work of objectivation. In other terms, the neuro-materialist strategy of explanation is determined by the very stance that had formerly created the “hard problem of the physical origin of conscious experience,” and it therefore has no capacity to dissolve it.

This confirms that a fully convincing dissolution of this problem requires a complete change of stance, as indicated by Varela: defocusing from intentional directness, balancing efforts between the first-person and the third-person approaches, and working actively to refining the neuro-experiential correlation rather than distorting it by way of a fake one-way explanation. In the same way as a Zen practitioner, the Varelian neurophenomenologist does not strive towards some solution to a standard problem. He rather exercises a “living and continuous reaction” (Batchelor 2000) that makes such a problem irrelevant.

## Acknowledgements

This article is based on a talk presented on 28 May 2011 at the conference “The Ethical Meaning of Francisco Varela’s Thought” (Sassari, Italy) in honor of the 10th anniversary of Francisco J. Varela’s death, organized by Rossella Mascolo.

## References

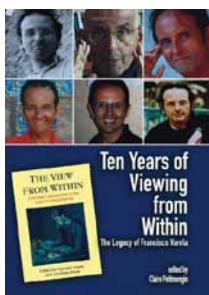
- Amstrong D. M. (1993) A materialist theory of the mind. Routledge, London.
- Batchelor S. (2000) Verses from the centre. Riverhead Hardcover, New York.
- Bayne T. (2004) Closing the gap? Some questions for neurophenomenology. *Phenomenology and the Cognitive Sciences* 3: 349–364.
- Bilodeau D. J. (1996) Physics, machines and the hard problem. *Journal of Consciousness Studies* 3: 386–401.
- Bitbol M. (2000) *Physique et philosophie de l’esprit*. Flammarion, Paris.
- Bitbol M. (2002) Science as if situation mattered. *Phenomenology and the Cognitive Science* 1: 181–224.
- Bitbol M. (2006) Une science de la conscience équitale: L’actualité de la neurophénoménologie de Francisco Varela. *Intellectica* 43: 135–157.
- Bitbol M. (2007) Ontology, matter and emergence. *Phenomenology and the Cognitive Science* 6: 293–307.
- Bitbol M. (2008a) Is consciousness primary? *NeuroQuantology* 6: 53–71.
- Bitbol M. (2008b) Consciousness, situations, and the measurement problem of quantum mechanics. *NeuroQuantology* 6: 203–213.
- Bitbol-Hespériès A. (2000) Descartes face à la mélancolie de la Princesse Elisabeth. In: Melkevik B. & Narbonne J.-M. (eds.) *Une philosophie dans l’histoire, hommages à Raymond Klibansky*. Presses de l’Université de Laval, Montréal: 229–250.
- Carnap R. (1936) *Existe-t-il des prémisses de la science qui soient incontrôlables?* Translated by Henri Buriot-Darsiles. *Scientia LX*: 129–135. Note: Carnap’s original German paper was never published.
- Chalmers D. J. (1996) *The conscious mind. In search of a fundamental theory*. Oxford: Oxford University Press.
- Chalmers D. J. (1997) Moving forward on the problem of consciousness. *Journal of Consciousness studies* 4: 3–46.
- Damasio A. (1995) *Descartes’ error: Emotion, reason and the human brain*. New York: Harper Perennial.
- Davidson D. (1980) *Essays on actions and events*. Clarendon Press, Oxford.
- Descartes R. (1991) *The philosophical writings III*. Cambridge University Press, Cambridge.
- Elie M. (1993) *Lumière, couleurs et nature*. Vrin, Paris.
- Feigl H. (1970) Mind-Body, not a pseudo-problem. In: Borst C. V. (ed.) *The mind-brain identity theory*. McMillan, London: 33–42.
- Gallagher S. & Zahavi D. (2008) *The phenomenological mind*. Routledge: New York.
- Hardcastle V. G. (1996) The why of consciousness: a non-issue for materialists. *Journal of Consciousness Studies* 3: 7–13.
- Henry M. (1985) *Généalogie de la psychanalyse*. Presses Universitaires de France, Paris.
- Husserl E. (1964) *The phenomenology of internal time-consciousness*. Translated by James S. Churchill. Indiana University Press, Bloomington. German original: Husserl E.



- (1928) Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins. Niemeyer, Halle.
- Husserl E. (1977) Cartesian meditations: An introduction to phenomenology. Martinus Nijhoff, Den Haag.
- Hut P. & Shepard R. N. (1998) Turning the "hard problem" upside down and sideways. In: Shear J. (ed.) Explaining consciousness, the hard problem. MIT Press, Cambridge.
- James W. (2003) Essays of radical empiricism. Dover Publications: London.
- Lachaux J.-P. (2011) If no control, then what? Journal of Consciousness Studies 18: 162–166.
- Lockwood M. (1993) Dennett's mind. Inquiry 36: 59–72.
- Lutz A. & Thompson E. (2003) Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. Journal of Consciousness Studies 10: 31–52.
- Merleau-Ponty M. (1964) Le visible et l'invisible. Gallimard, Paris.
- Natorp P. (1912) Allgemeine Psychologie nach kritischer Methode. Tübingen: Mohr.
- Peschard I. & Bitbol M. (2008) Heat, temperature and phenomenal concepts. In: Wright E. (ed.) The case for qualia. MIT Press, Cambridge: 155–174.
- Petitmengin C., Navarro V. & Baulac M. (2006) Seizure anticipation: Are neurophenomenological approaches able to detect preictal symptoms? Epilepsy and Behavior 9: 298–306.
- Petitmengin C. & Bitbol M. (2009) The validity of first-person descriptions as authenticity and coherence. Journal of Consciousness Studies 16: 363–404.
- Place U. T. (1956) Is Consciousness a brain process? British Journal of Psychology 47: 44–50.
- Rudrauf D., Lutz A., Cosmelli D., Lachaux J.-P. & Le Van Quyen M. (2003) From autopoiesis to neurophenomenology: Francisco Varela's exploration of the biophysics of being. Biological Research 36: 27–65.
- Ryle G. (1949) The concept of mind. The University of Chicago Press, Chicago.
- Schrödinger E. (1967) What is life? Mind and matter. Cambridge University Press, Cambridge.
- Searle J. (1992) The rediscovery of the mind. Cambridge University Press, Cambridge.
- Taraborelli D. & Mossio M. (2008) On the relation between the enactive and the sensorimotor approach to perception, Consciousness and Cognition 17(4): 1343–1344.
- Thompson E. (2007) Mind in life. Belknap Press, Cambridge MA.
- Thompson E. & Varela F. J. (2001) Radical embodiment: Neural dynamics and consciousness. Trends in cognitive science 5: 418–425.
- Van Fraassen B. (1980) The scientific image. Oxford University Press, Oxford.
- Varela F. J. (1976) Not one, not two. The coevolution quarterly. Fall 1976: 62–67.
- Varela F. J. (1996) Neurophenomenology. Journal of Consciousness Studies 3: 330–349.
- Varela F. J. (1997) The naturalization of phenomenology as the transcendence of nature: Searching for mutual generative constraints. Alter: Revue de Phénoménologie 5: 355–385.
- Varela F. J. (1998) A science of consciousness as if experience mattered. In: Hameroff S., Kaszniak A. W. & A. C. Scott (eds.) Towards a science of consciousness II: The second Tucson discussions and debates. MIT Press, Cambridge: 31–44.
- Varela F. J. & Shear J. (1999) The view from within. Imprint Academic, London.
- Varela F. J. (1999a) Dasein's brain: Phenomenology meets cognitive science. In: Aerts D. (ed.) Einstein meets Magritte. The white book. Kluwer, Dordrecht: 185–197.
- Varela F. J. (1999b) Present-time consciousness. Journal of Consciousness Studies 6(2–3): 111–140.
- Varela F. J. & Depraz N. (2000) At the source of time: Valence and the constitutional dynamics of affect. Journal of Consciousness Studies 12: 64–81.

RECEIVED: 5 JUNE 2011

ACCEPTED: 16 APRIL 2012



### OF RELATED INTEREST TEN YEARS OF VIEWING FROM WITHIN

The book *Ten Years of Viewing from Within: The Legacy of Francisco Varela* edited by Claire Petitmengin is a collection of essays that examine and refine the research program on first-person methods with contributions based on empirical research. For an extended review see Tom Froese, Cassandra Gould and Adam Barrett (2011) Re-Viewing from Within: A Commentary on First- and Second-Person Methods in the Science of Consciousness. Constructivist Foundations 6(2): 254–269, available at <http://www.univie.ac.at/constructivism/journal/6/2/254.froese> Imprint Academic, Exeter, 2009. ISBN 978-1845401740. 256 pages.