On the radical self-referentiality of consciousness

Michel Bitbol
CNRS / Ecole Normale Supérieure (Archives Husserl), Paris

The present text (presented in some conferences) picks out the philosophical nucleus of M. Bitbol & P.-L. Luisi, "Science and the self-referentiality of consciousness", *Journal of Cosmology,* 14, 4728-4743, 2011. It contains a short outline of a more intricate argument to be found in chapter 1 of : M. Bitbol, *La conscience a-t-elle une origine?,* Flammarion, 2014

Abstract
This paper is concerned by the extent and limits of scientific inquiry about consciousness. I first insist on the exceptional status of conscious experience, which is no proper object of investigation, but rather an actual presence and a precondition of any investigation. To better characterize this status, I develop the concept of "radical self-reference". Questioning about consciousness is radically self-referential in so far as it is itself an act of consciousness. This suggests that consciousness is *existentially* primary; a kind of primacy which clearly departs from the *ontological* primacy advocated by property dualists or panpsychists. I then notice that, accordingly, science has some basic features which hinder in principle its approach of consciousness: it distantiates from its objet, whereas consciousness is at no distance; it tends to formulate truths that do not depend on one's situation, whereas consciousness is what it is like to be situated; it claims that physical explanations are self-sufficient, thus threatening to reduce consciousness to an epiphenomenon. These remarks tend to increase the "hardness" of the "hard problem" of the *origin and existence* of consciousness. Yet, I also point out that scientific inquiries are able to clarify a host of interesting issues about the *forms and development* of consciousness.

According to Maurice Merleau-Ponty (1964), philosophy is "(…) the set of questions in which the one who questions is himself implicated in the question". Any question about consciousness is then utterly philosophical. For when we raise a question about consciousness, we are not only implicated in it *in abstracto*, timelessly, as generic human beings ; we are fully implicated in it by what we *are* at this precise moment. We are fully and presently implicated because *formulating* a question about consciousness is an act of consciousness ; *understanding* a question about consciousness is an act of consciousness ; figuring out how we could *answer* a question about consciousness is yet another act of consciousness. In short, questions about consciousness are radically self-referential.

Let us ponder about this notion of *radical* self-referentiality, because it may bring us closer to the heart of the issue of consciousness. A sentence such as “this sentence uses five words” is self-referential. It is easy to see that it indeed refers to itself, provided one shifts attention from the meaning of the sentence to its lexicon, from what it says to the graphemes it is made of. Here, the required attention shift goes from one object of consciousness (what is meant) to another object of consciousness (the written words). By contrast, the self-referential character of a question about consciousness is seen only if attention shifts from the meaning of the question to present conscious experience as the background of this very act of attention. In this case, the second focus of attention is no “object” at all ; rather, it is the condition for anything to be taken as an object. This is why questions about consciousness are more than self-referential : they are *radically* self-referential.

The standard question “*where* does consciousness come from ?” provides us with a good illustration of how misguided one can be if this radical self-referentiality is ignored. When we ask the question “where ?”, we prepare ourselves to focus our attention on some restricted region of our conscious experience : right or left, up or down, nearby or far away, inside or outside the skull, in this or that part of the brain. And when we think we have got the answer, after a deep speculative reflection or after a long experimental inquiry, this answer inevitably consists in pointing towards an *object* or a process that we can describe, think about, or even sometimes *imagine*. In other terms, answering a question about the origin of consciousness is tantamount to singling out a given *content* of our consciousness, and encouraging others to modulate their own consciousness accordingly. Everything looks as if we were trying to ascribe consciousness as a *whole* to some *part* of it ; as if conscious experience, this all-pervasive fact that constitutes our lives, were tentatively encapsulated in a *fraction* of it. This sounds awkward indeed !

There is an easy way to alleviate this feeling of awkwardness, though. Turning our attention to the background condition of any act of attention (in line with radical self-referentiality), we are bound to reply that “consciousness comes from nowhere else than

… *here*". True, "here" does not look like a serious answer, because it does not refer to a special place, a special object, or a special process. But should we dismiss it so quickly ? Let us think a little further. "Here" is an indexical (or deictic, or demonstrative) term, like "I", "now", and "this". As any indexical term, it fully *commits* the person who utters it. It thereby invites other persons to figure out how things appear from the standpoint of the utterer, or more generally how things appear from the standpoint of any utterer whatsoever. "Here" is a verbal operator that brings each one of us back to one's own situation. Saying that "consciousness comes from *here*" then means that consciousness has no other obvious origin than the actual situation. Consciousness is the name we give to the astounding realization of immediate existence, even before its more intricate connotations such as reflective self-consciousness or moral conscience. Consciousness, in this very elementary sense, is *existentially primary*.

These obvious (yet destabilizing) remarks are not derived from any reasoning. They rather arise when we suspend any judgment, and just state the elementary features of what we are living. They express what E. Husserl (1913/1931) called a *phenomenological description ;* a plain statement of what is immediately experienced, irrespective of any interpretation of the contents of experience in naturalistic terms. So, asserting that consciousness is "existentially primary" is no metaphysical doctrine ; this is no idealist or panpsychist doctrine of the ontological primacy of consciousness to be contrasted with a doctrine of the ontological primacy of matter. This is just an invitation to be faithful to our own lived experience in its most pristine form.

Is such lack of reasoning a defect of the (phenomenological) approach ? Actually, it might well be its major quality. Indeed, as E. Schrödinger (1964, p. 19) noticed, when the problems of mind and consciousness are dealt with, *the reasoning is part of the overall phenomenon to be explained, not a tool for any genuine explanation*. Here again, radical self-referentiality must be taken into account. As any reasoning, a reasoning about consciousness involves a conscious experience ; *aknowledging the validity* of a personal reasoning, or even of a mechanical inference performed by a Turing machine, is still a conscious experience. A reasoning

bearing on consciousness is included in what is reasoned about. So, when consciousness is presented as an object of reasoning, this can only be in a fake sense.

In fact, as soon as we embark on anything like discourse, reasoning, or scientific research about consciousness, we are driven away from mere aknowledgment of what is lived now, and thereby away from the central topic of the inquiry. So much so that recovering contact with it becomes difficult, and that, from then on, we tend to value more the abstract product of arguments than their experienced source.

Let us first ponder about discourse and language. Language *means* and *discriminates*.

Meaning is tantamount to displacing attention. It displaces attention from the sound of a word to what it signifies, from the pointed finger to what it aims at showing. Meaning thereby pushes us outwards, towards the future, towards something that is *not* close at hand. When we use a *word* for "consciousness", we are then automatically led astray, because conscious experience is not some*thing* over *there* to be *meant* in any way. Once again consciousness is plainly *here* ; this "here" that submerges us ; this "here" that is presupposed by any location in space. Trying to *mean* consciousness is self-defeating, since what is allegedly meant is nothing beyond the very act of meaning it. It is radically self-referring.

The same holds for the discriminative power of language. How can we discriminate present conscious experience from anything else ? Should we discriminate it from brute matter ? But brute matter is only given or thought now *qua* object of conscious experience ! Should we discriminate it from its absence at certain moments of our lives (such as sleep or fainting) ? But these moments are only known now *qua* contents of present conscious experience ! At this point, we are ready to understand some cryptic remarks in which Wittgenstein speaks of consciousness "… as the very essence of experience, the appearance of the world, the *world*". Consciousness is coextensive to the world, because no world has ever been given independently of one's conscious realization of it. In the wake of this remark, Wittgenstein points out that "if I had to add the *world* to my language it would have to be

one sign for the whole of language, which sign could therefore be left out" (Wittgenstein, 1982, p. 42). In view of the equivalence between consciousness and world, the same thing can be said of consciousness. By using a word for "consciousness", we try to discriminate it from something else. This can be done in everyday use for making a difference between somebody else's *apparent* states of wakefulness and sleep ; but not in the proper existential (and radically self-referential) sense, since *at this precise moment* that contains in it all the memories of the past and all the projects for the future, there is nothing that can be contrasted with it.

And what about science ? What about the physics that physicalist doctrines of consciousness refer to ; what about the neurobiology that reductionism or eliminativism put forward ? To begin with, science uses language, symbols, and reasonings. Science is therefore biased about consciousness in the same way as language itself : it attempts to distantiate what is at no distance from us, and discriminate what can be contrasted with nothing. Drawing from language, scientists tend to treat consciousness as a *property* of human organisms. However, they should know that ascribing a "property" to something must be based on reliable *criteria* bearing on this thing ; whereas any bodily consciousness-criterion, be it presence or absence of verbal report, or presence or absence of certain waves on an electroencephalogram, is weak and ambiguous. The only true evidence, the only absolute criterion, is *first-personal*. Elementary consciousness, pure experience, is thus no "thing" and no "property" ; it is *an all-pervasive precondition for referring to things and properties* (Bitbol, 2000, 2008, 2014). But this is only part of the difficulty. More specifically, science was born from the decision to *objectify*, namely to select the elements of experience that are invariant across persons and situations. Its aim is to formulate *universal* truths, namely truths that can be accepted by anyone irrespective of one's situation. Therefrom, the kind of truths science can reach is quite peculiar : they take the form of universal and necessary *connections* between phenomena (the so-called scientific *laws)*. This epistemological remark has devastating consequences. It means that *in virtue of the very methodological presupposition on which it is based*, science has and *can* have *nothing* to say about the mere fact that *there are*

phenomena (namely appearances) for anybody, let alone on the *qualitative* content of these phenomena (Wright, 2008).

Let us give a few examples. Physics establishes laws about electromagnetic phenomena. It classifies the waves that give rise to the perception of colors according to their wavelengths. But it has nothing to say about the very existence of an experience of color and even less about its lived quality. Psychophysics and neurology of occipital cortex areas add more and more precise knowledge about the *structure* of color perception in humans, about the mutual *relations* of various perceived colors, and about the physiological states in which color perception is reported to be altered. But these sciences remain mute about how and why there should *be* any lived experience of color at all when neuronal activity occurs in these brain areas, and about *what it is like* to experience blue or red. More generally, we have witnessed amazing advances of neurophysiology about how the brain stores information, binds its maps and programs for action, and even elaborates self-mapping. These discoveries have also been carefully *correlated* to human subjects' descriptions of their own conscious experience, thus allowing scientists to speak of memory instead of information storage, of unified consciousness instead of information binding, and of self-awareness instead of self-mapping. But nothing, not the slightest clue, has been provided about why and how these neuronal processes should generate anything like conscious experience. In other terms, borrowed from David Chalmers, physical and neurological sciences have shown their ability to solve an unlimited number of "easy problems" about the structure and neural correlates of conscious events, but they remain silent about the "hard problem" of the existence, origin, and "feel" of conscious experience itself.

This is no defect of science, nor is it a temporary obstacle that one may hope to overcome in some remote future. This is just a consequence of the methodological decision to *objectify* that has been taken at the very foundation of science. Objectification automatically pushes situatedness and lived experience in the "blind spot" of research. *No amount of scientific effort can recover what has been lost by basing science on such principle.* Some authors [e.g. Hardcastle, 1996] have then argued that science

should be allowed to completely *ignore* the "hard problem" and just proceed with the many interesting "easy" problems it is able to clarify.

There are symptoms in the philosophy of mind showing that this fundamental limitation of the scientific inquiry about consciousness is not taken seriously enough. One of them is the so-called "causal closure" of physical and physiological explanations. Nothing prevents one from offering a *purely physical or physiological* account of the chain of causes occurring from a sensory input received by an organism to the behavior of this organism. At no point does one need to invoke the fact that this organism is perceiving and acting *consciously*, that it has a *feel*. The same is true of evolutionist arguments. Evolution can select certain useful *functions* ascribed to consciousness (such as unification of information, or behavioral emotivity of the organism), but not the mere fact that *there is something it is like* to implement these functions. In other terms, borrowed from N. Block et al. (1997), evolution can select adaptative features of *access* consciousness but not the presence of *phenomenal* consciousness itself. An interesting application of the evolutionary argument to some functional aspects of consciousness will be documented later. But at this point, we must aknowledge that in any *fully consistent* scientific account, phenomenal consciousness is bound to be causally irrelevant or *epiphenomenal*.

And yet, several neurological theories of consciousness have been formulated in the last few years, with some success. One of them is the *global workspace theory*, according to which consciousness arises when information is retrieved from several specialized modules of the brain, and then assembled in a broadly distributed neural network involving a central working memory. It remarkably accounts for some facts of experience, such as the famous "binocular rivalry" reported by subjects who have been presented different images to each eye. Another theory is the *integrated information theory*, according to which consciousness arises when the neural processes are both rich of information and highly cross-linked in time and space. This theory is remarkably helpful because it endows the structure of electromagnetic signals of the brain with a power of *prediction* of future reports of

conscious experience in patients with coma, vegetative states, or general anesthesia. However, once again, none of these theories clarifies the *origin* of consciousness. As any scientific theory, they systematically connect and predict phenomena. They connect neurological phenomena with behavioral phenomena such as verbal reports of patients. But they do not offer the smallest hint to explain the elementary fact that there *are* phenomena, that there *are* appearances at all.

One may protest at this point, by adducing a more common-sense argument in favor of neurophysiological reductionism. It is obvious that the eyes are organs of vision, because when we close our eyelids we see no longer. In the same way it should be obvious that the brain *is* the organ of consciousness, because if it is switched off, no consciousness is left. Even if we do not know the precise process, consciousness is then bound to come from our brain. But, before we jump to hasty reductionist conclusions, let us investigate this notion of "switching off" the brain in more details. General anesthesia can be used to modulate the process of switching off, because one may control the amount of drugs. When doses of drugs are increased, one observes that the functions which are usually merged into a single concept of consciousness are lost not together, but in succession. One successively loses explicit long-term memory, ability to report verbally, social intercourse, coordinated behaviour, implicit memory, wakefulness etc. Now, the asymptote of this series of losses is strictly unknown. When the brain is entirely switched off, with flat EEG, is there any experience left, or not ? One answer is that there is nothing left indeed. But another answer is just as compatible with the data we know. This alternative answer is that switching off the brain only abolishes the discriminative, cumulative, narrative, self-monitoring, unifying functions of consciousness ; yet it retains instantaneous, un-discriminative, un-cumulative, un-self-monitored experience : *A sort of blank, forgetful and contentless experience*. In this latter case, the power of the brain would not be to generate consciousness *ex nihilo,* but only to bind, focus, accumulate, and bring to self-reflection the all-pervasively *given* experience within a coherent situated knowledge. *These* latter functions are in principle accessible to science ; they are part of the so-called "easy

problems". In particular, one may adduce very interesting evolutionary arguments in order to explain the emergence of crucial features of our human consciousness, such as *self-reflection* (which is too often identified with consciousness *in general* along with scientific discussions). It may well be the case that becoming intellectually and emotionally aware of ourselves, of our being distinct from the environment, of our finiteness in time that manifests by death through self-reflection, represented a remarkable behavioral or reproductive advantage. It may well be the case that cristallizing the awe about this existential situation in rituals and spiritual practices rendered human tribes more cohesive and more united, and that this proved crucial for survival. It may also be the case that these new features of self-awareness and emotive existential realization have been favored by certain genetic alterations that were selected thereby. But these relevant evolutionary theories must not be mixed up with a true reductionist scientific account of the radical origin of conscious experience. Indeed, they can easily be understood in terms of a non-reductionist conception of the relation between the lived and the living, between conscious experience and biological processes. One such conception was remarkably expressed by Beauregard and O'leary (2007) : "More than a century ago, William James proposed that the brain may serve as a permissive / transmissive / expressive function rather than a productive one, in terms of the mental events and experiences it *allows* (just as a prism, which is not the source of the light, changes the incoming light to form the colored spectrum). Following James, Bergson and Huxley speculated that the brain acts as a filter or reducing valve by blocking out much of, and allowing registration and expression of only a narrow band of, perceivable reality. They believed that over the course of evolution, the brain has been trained to eliminate most of those perceptions that do not directly aid our everyday survival. This outlook implies that the brain normally *limits* the human capacity to have spiritual experiences". Such vision of the brain and body as *modulators* and restrictive *filters* makes sense of scientific theories and experiments about the many-layered functionalities and amplifications of consciousness, without underrating the significance of the fact that raw experience is our

most immediate and basic given. It helps us to gain a better understanding of another true wonder : that there is not only "something it is like to be", but that there is also *reflective realization* of being.

To conclude, we must come back to where we started. These reflections on consciousness were (and were bound to be) mostly philosophical. This may be taken by some as a defect ; for, as anybody knows, philosophy is sterile. But is it really so ? As Heidegger (1935/1998) noticed, *we cannot make anything with philosophy, but philosophy can make something of us*. Philosophy does not change the world, but it can change us, our outlook, our ways, our minds. A conception of consciousness relying more on how we live it than on how we can manipulate its contents, could change a lot in our (medical and non-medical) way of approaching our fellow human beings. It could promote sharing, empathy, and clinical patience, by giving our "transexperiental"[1] intercourse its full credentials.

Bibliography

Beauregard M. and O'Leary D. (2007), *The spiritual brain,* London : HarperOne

Bitbol M. (2000), *Physique et philosophie de l'esprit,* Paris : Flammarion

Bitbol M. (2008), "Is Consciousness Primary ?", *NeuroQuantology,* 6, 53-71

Bitbol M. (2014), *La conscience a-t-elle une origine?,* Paris : Flammarion

Block N., Flanagan O., Güzeldere G. (eds.) (1997), *The Nature of Consciousness,* Cambridge : M.I.T Press

Hardcastle V. (1996), "The why of consciousness : a non-issue for materialists", *Journal of Consciousness Studies,* 3, 7-13

Heidegger M. (1935/1998), *Einfuehrung in die Metaphysik,* Berlin : Niemeyer

Husserl E. (1913/1931), *Ideas: General Introduction to Pure Phenomenology,* London: George Allen & Unwin

Merleau-Ponty M. (1964), *Le visible et l'invisible*, Paris : Gallimard

Schrödinger E. (1964), *My view of the world,* Cambridge : Cambridge University Press

Wittgenstein L. (1982), *Notes on private experience and sense-data,* Paris : T.E.R

Wright E. (2008), *The Case for Qualia,* Cambridge : MIT Press

[1] This is a word Ronald Laing often used to denote our unformulated presupposition that the other has a lived experience which can be shared with us in dialogue.