About Metzinger's ego tunnel

by Astro Calisi

Thomas Metzinger is without doubt one of the most prominent representatives of contemporary philosophy of mind in Germany. His thesis are original and have as their starting point the most peculiar features of consciousness, such as they present themselves to us. He recognizes the consciousness a central role in managing behavior of living organisms, and indeed he believes that only with due care considering the richness and depth of conscious events, without preconceptions and fears of any kind, we can get to understand the real nature of this faculty. (1)

In this short paper I propose to consider in its essential Metzinger's theory of consciousness, highlighting some key points compared to other positions in current debate on mind. At the same time, I can't avoid to expose the limits of this theory, showing how it ultimately rests on overly casual use of terms and concepts whose location within the scientific worldview is to consider problematic, at least in the present state of our knowledge.

I shall not deal, however, with ethical considerations that Metzinger develops abundantly in the last three chapters of his book. I think these considerations, although presented as a consequence of the possibility of creating an artificial system endowed with consciousness (a "machine ego"), have in fact the function of strengthening plausibility of the proposed overall elaborated by Metzinger.

Metzinger's concept of consciousness has as its immediate precursors Antonio Damasio and Bernard Baars. These authors, although from different perspectives, had already highlighted the essential role of consciousness in making the information available to the body in a global and strongly unified shape (2). Metzinger goes ahead. In his proposal consciousness contents and the self that everyone feels – or rather, *lives* – as the center of his being are nothing but a representation built by the brain to facilitate the interaction of organism with the environment (3). Subjective experience should therefore be seen as a very specific way of organizing the information about outside world, in the form of a "knowledge of self": a knowledge *in the first person*. (4)

An important precondition for such a representation takes place is that the subject is not aware of the processes by which it is carried out. Metzinger insists much on this point. In fact, it is a bit the cornerstone of the entire theory proposed by him: to build up in the form of conscious experience of someone, in order to have a center, identified as *I*, the representation must be available to the subject as the reality of the world and not as an image built from that reality. Metzinger coined the term *ego tunnel* to indicate that consciousness model is *transparent* to itself, meaning by this that the subject is not aware of the ways in which information is processed to take the form in which it is presented to him (5). Indeed, the same subject must be considered a representation which, in turn, does not know to be so. (6)

According to this view, the representation of reality as conscious experience and the image that the system has got about himself would be a construction processed by the brain in order to allow our body to see himself as a highly integrated entity, thus facilitating planning and control of behavior. (7)

The assumption that lies at the base of such a conception is that a brain became extremely complex problems to manage in the best way its functions without reference to an overall image which includes both the information available at a given time, and the activities to be undertaken or

in progress. I will express in a little while about the sustainability of this assumption and especially the idea that a representation, however it is conceived, may increase somewhat efficiency of any system. For now, I just want to anticipate that this thesis seems to be based on a kind of conceptual mistake, even if hardly recognizable as such.

I think, however, it is advisable to preface my criticism a consideration of the positive contents of Metzinger's proposal; in fact, they put to our attention aspects and implications of consciousness routinely underestimated or completely ignored. Metzinger shows to take seriously consciousness, highlighting some important features, such as its being related lo less stereotyped and more creative behaviors (8); its ability to present its contents as *experiences of someone*, thus promoting integration of different types of information available at a given time (9); and its selective, ie its being directed only to aspects of reality which are relevant to a possible action to be undertaken (10).

About the latter feature, it may be added that not only conscious attention takes into consideration just the elements that are significant from a behavioral point of view, but it also tends to present them in the best suited form to be used in action, in order to reduce to a minimum the need for further processing at a conscious level. For example, noises and sounds come to us already placed in the source that emits them, and not in our ears where they are actually detected; even visual objects you have already placed in outside world, and, moreover, in their correct direction. While, we know that the projected image on retina is inverted compared to reality.

We are dealing, as you easily realize, with mechanisms which, by reducing the commitment of consciousness on repetitive tasks, make this available for more demanding tasks, usually related to greater behavioral plasticity.

It is quite clear that consciousness plays an adaptive role of major importance in living organisms behavior. This recognition of Metzinger is - in my opinion - one of the main strengths of his proposal. While many authors not only give no efficacy to consciousness, but take the too easy strategy to solve the problems posed by it through an operation of a substantial devaluation, which can sometimes go to extremes of total denial. (11)

Consciousness model presented by Metzinger seems at first sight plausible because it closely resembles our immediate experience. It talks about us, how we live what appears moment by moment in the domain of our subjectivity, giving an interpretation of it that is broadly consistent with what we know about our brain.

The *ego tunnel* is the key concept of this explanation: we are not able to track the brain processes to conscious experiences as the brain processes that create subjective experience are inaccessible to our consciousness. Indeed, this inaccessibility is one of the basic conditions for such an experience takes consistency in its characteristic shape. Even the ego, personified entity to which everything is referred to, is a brain product - a virtual no consistency in the real - which knows nothing about its origin. The ego and consciousness are transparent to themselves, and that transparency allows them to form as factors that contribute to adaptation of organisms to the environment.

However Metzinger doesn't tell us anything about the main problem concerning the relationship between mind and brain; he doesn't address the crucial issue of how it makes possible conscious experience, a firsthand experience, originated by impersonal physical phenomena such as those that occur in the nervous system. He focuses in an effort to explain why conscious experience and the ego present themselves as *something other* than cerebral activity from which they derive. In doing so, however, he incurs a real mistake on a conceptual level, putting together undue physical phenomena (nerve activity of the brain) with other phenomena and properties typically "mental", which is very difficult to settle completely to the previous field of phenomena .

To fully understand how questionable is the attempted operation by Metzinger, let us try to immerse ourselves in his proposal, taking as a reference not our conscious experience – the subjective experiences – but the nervous organization of the brain or the organization of an information processor. In fact, these are the "systems" that we have to face when we want to move in a truly "naturalistic" perspective.

How can we frame within these systems a concept such as *transparency*? How could it be achieved (or simply imagined) transparency in a system acting in a totally impersonal way, based on well-defined scientific laws?

In a similar vein, transparency can only occur under some form of *inaccessibility*: inaccessibility of specific information contents to the system, or maybe to a part of it. According to this view, the brain – as well as a suitably programmed computer – would generate an image of itself and about its operating way, and these ones would have no access to information about their own constitution. What could this mean in terms of functionality? How should we represent the idea of an image that "ignores" the processes that led to assume its present form? Or better: is it conceivable an image capable of any self-representation?

Let us ask, first, of which nature may be an image produced by an artificial system, or a brain, the operations of which are strictly based on exchange of electrical (or electro-chemical) signals between the constituent units. Despite our efforts, it is unlikely that we can represent it differently from one of the following forms: *visual*, a set of bulbs or LEDs that turn on and off according to specific sequences or spatial configurations, *signs* engraved on any media, or *graphs* of some kind, which appear on a monitor, *sound signals*, also produced according to some logic to translate information, or *maps* of data stored in the system memory. But, on closer inspection, it would be an unnecessary complication. Such images, in order to be used, should in any case be interpreted (decoded) from the same system (the brain or the computer) that previously had built them in that form. So what's the advantage over the behavior? What king of information would add the image constructed by the system (or a part thereof) that were not already available to the system itself? In short: why couldn't the (biological or artificial) system control the various functions, cognitive and motor skills, in a completely mechanical and impersonal way, using the whole wealth of information in its possession, as they do – without exception – all machines so far man-made?

From a purely informational point of view, affirming – as Metzinger does (like Damasio and Baars) – that bringing together all the relevant information in one virtual place is advantageous for action to be undertaken, means acknowledging information in this way organized a higher content compared to the same information when it is spread out in various areas of the same system. But *who* (or *what*) is organizing information in this form, if not the same system? Since the information can't be created from nothing, and any transformation of it from one form to another requires further information, we are forced to conclude that the apparent increase of the information content must correspond exactly to the information used by the system to select and organize in a particular way the relevant information to a specific action.

All this makes even more pressing the question: why couldn't the system operate fully automatically, based on all the information available?

To such a question we have no answer within the consciousness perspective developed by Metzinger. In fact, using the concept of transparency (implicit in that of "ego tunnel"), he only assumes (or, better, *postulates*) the existence of a conscience, rather than explain it. Philosophical concept of transparency, on a closer inspection, is built by taking as a reference the features of our thinking, and in particular the feature to reflect on itself. If we eliminate the ability of self-reflection – typical mental capacity – consciousness disappears too: the transparency can only result in a much more modest inability to access certain information. But, as we have seen, the

inaccessibility so conceived becomes a pure non-sense when applied to systems operating in accordance with ordinary physical principles.

Thus we see that consciousness model focused on the "ego tunnel" is not able to tell us anything relevant to the arising of consciousness. In fact, it takes for granted the *existence* of consciousness when it introduces as a non-problematic the brain representations in the form of subjective experience. The same happens when it uses the concept of *transparency* to justify our inability to retrace the steps which lead from nervous processes of the brain to our subjective experience. Transparency can't be realized outside a context in which it is potentially present some conscious capacity.

This basic misunderstanding is the price Metzinger has to pay to make, at least in appearance, his theoretical perspective compatible with the scientific conception of world. It is, moreover, an explanatory strategy much more common than what we might think: it seems indeed that we can't help but resort to some kind of trick on a conceptual level every time we deal with the problems of mind-brain relationship within the consolidates categories of science. Devices can be more or less sophisticated, more or less difficult to identify as such, but they always has the function to divert our attention from the most difficult aspects to be addressed, or to transform these aspects in apparently not-problematic and at first sight acceptable forms. (12)

However, it is not with this kind of tricks we can hope to overcome current problems posed by the mind. It's true that, at the time, no one is able to suggest solutions which are really satisfactory, but at least we can get to work so that these deceptive strategies will be exposed for what they are, so that fewer and fewer people waste time with them.

NOTES

(1) Thomas Metzinger, *The Ego Tunnel: The Science of the Mind and the Myth of the Self*, Basic Books, New York, 2009; Ital. transl., *Il tunnel dell'io. Scienze della mente e mito del soggetto*, Raffaello Cortina, Milano, 2010, p. 2.

(2) Antonio Damasio, *Alla ricerca di Spinoza*, Adelphi, Milano, 2003; Bernard Baars, *A Cognitive Theory of Consciousness*, Cambridge University Press, New York, 1988.

(3) Thomas Metzinger, *Quot. work*, p. 4.

(4) *Q. w.*, p. 9.

(5) *Q. w.*, p. 8.

(6) *Q. w.*, pp. 8-9.

(7) *Q. w.*, pp. 4-5.

(8) *Q. w.*, p. 66; see also p. 233.

(9) *Q. w.*, pp. 67-70.

(10) *Q. w.*, p. 7.

(11) This tradition of thought, very roughly, can be traced from *The Concept of Mind* (1949) by Gilbert Ryle. It ideally goes on with Daniel Dennett, till the recent writings by Susan Blackmore. For a criticism to Dennett's "eliminatist" position, see Astro Calisi, *Oltre gli orizzonti del conosciuto. La sfida cruciale della mente alla scienza del XXI secolo*, Uni Service, Trento, 2011, pp. 73-87. With regard to Blackmore's thesis, I tried to show their weakness in Astro Calisi, "La prospettiva della mente di Susan Blackmore", on Psicolab: http://www.psicolab.net/2011/prospettiva-mente-susan-blackmore .

(12) Astro Calisi, *Oltre gli orizzonti del conosciuto*…, Quot., pp. 231-270.

[From Astro Calisi, "Il tunnel dell'io",
in http://www.sitosophia.org/recensioni/il-tunnel-dellio/.
English translation from Italian by the autor – sept. 2012]