

Physical Foundations of Phenomenal Content in a Conscious Machine

Riccardo Manzotti

*Institute of Consumption, Communication and Behaviour
IULM University, Milan, Via Carlo Bo, 8 20143 – Milano
E-Mail: riccardo.manzotti@iulm.it*

Abstract

This paper outlines a series of very short statements which sketch the backbone of a hopefully coherent framework that could ground future research on artificial consciousness. The paper aims at being very provocative. It tries to convey the intuition that we need to tackle with foundational issues lying underneath many commonly accepted assumptions – something along the spirit of Albert Einstein's words (quoted in Jammer, 1954/1993, p. xiii-xiv):

In the interest of science it is necessary over and over again to engage in the critique of fundamental concepts, in order that we may not unconsciously be ruled by them. This becomes evident especially in those situations involving development of ideas in which the consistent use of the traditional fundamental concepts leads us to paradoxes difficult to resolve.

It appears that consciousness belongs to this kind of situations. The paper is divided in two sections – the former with a set of claims on classic issues about natural and artificial consciousness and the latter about a set of crucial issues which are seldom mentioned.

1 Classic issues

1.1 Machine consciousness entails phenomenal consciousness

Although many scholars suggested that it is possible to distinguish between phenomenal consciousness and not-so-awkward form of consciousness (such as resilience, autonomy, access-consciousness, and many others), the core of the problem remains phenomenal consciousness. This is not to say that all the issues mentioned – autonomy for instance – are not worthy of further enquiry. On the contrary, they are often both extremely important and enormously difficult. Moreover, they are probably necessary in order to have a conscious being, as many other things (a body for instance). However, they are not what we mean for consciousness. To be conscious entails having an experience of what is taking place – namely, to feel it. This has been pointed out very firmly by Stevan Harnad (Harnad and Scherzer, 2008). Explaining consciousness is the same as explaining what is feeling something. So, we can argue that feeling is a certain phenomenon (whether we agree or not on that), but we cannot simply drop the issue either of feelings or of phenomenal consciousness in favor of something else easier to manage.

If explaining consciousness entails explaining phenomenal consciousness or feeling, machine consciousness entails designing and building machine capable of phenomenal consciousness or feeling.

In short, a conscious machine ought to be a phenomenally conscious machine.

1.2 Artificial and natural consciousness

In 2001, at the Cold Spring Harbour Laboratories sponsored by the Swartz Foundation, it was held a seminar on the topic whether machines could ever be conscious. It was concluded that “There is no known law of nature that forbids the existence of subjective feelings in artefacts designed or evolved by humans” (quoted in Aleksander, 2008).

In a similar way, there is no known law of nature that entails that neurons are both necessary and sufficient to produce phenomenal consciousness. Further, neurons are not special in any way apart from being a very smart kind of cell doing many tricks – none straightforwardly connected to consciousness. Of course, it cannot be ruled out that someone in the future will discover some intrinsic neural feature responsible for consciousness (something along the lines of unknown quantum phenomena or similar). However, up to now, nothing like it has been identified, even remotely.

This is good for artificial consciousness since it paves the way for a system endowed with phenomenal consciousness and not made of biological neurons.

1.3 Phenomenal consciousness poses an ontological problem

This point is basically the hard problem. However, there is a difference. David Chalmers's view rules out the possibility that the physical world had any quality although he later suggested some kind of property dualism – any kind of dualism meaning that there is the need of a twofold structure of reality. However, we are not completely sure about the features of the physical world. For instance, it could be argued that since we feel something this something should belong to the physical world (Eddington, 1929/1935; Skrbina, 2005; Strawson, 2006). In the past scientists and philosophers developed a model of the physical world which is not compatible with phenomenal content. As it happens, the model could be wrong and there could be space for phenomenal content in the physical world. Instead of desperately and repeatedly trying to reduce phenomenal content to a dogmatic quality-less world, the other strategy could be exploited – is there any space for quality in the physical world? Maybe we are committed to some kind of epistemic blind spot which prevented us from describing the physical world in its completeness.

1.4 Phenomenal consciousness is the result of natural selection

Is there really anything which is not? Even features, which are highly dubious in their usefulness like the *vernis caseosa*, are not there by chance. The epiphenomenal view has never been able to cope successfully with the overarching power of natural selection. Furthermore, phenomenal consciousness seems to be a rather ubiquitous feature of complex cognitive organisms and not a rare oddity like the *vernis caseosa*.

If this is the case, there are two possible options. Either phenomenal consciousness provides some advantages per se, which is dubious, or phenomenal consciousness is the expression of a certain structure which is advantageous. Either way, phenomenal consciousness will be useful to a machine.

1.5 Phenomenal consciousness cannot be a functional state

Many authors tried to reduce phenomenal consciousness to functional states. The general idea is that to be conscious of x is the same as being in some functional relation with x. This did not work for many reasons that have been done to death. In short, the argument can be synthesized in three main problems: i) functional states do not address qualitative content, ii) functional states do suffer of multiple potential phenomenal content, iii) functional states are epistemic entities and not real stuff.

1.6 Phenomenal content cannot be a re-presentation

Re-presentations are arbitrary physical structure that can be used for functional purposes in machines or in human interactions. However they are different from what they re-present and do not have any autonomous re-presentational power as such. They are called re-presentation for historical and practical reasons that should not lead to a misleading metaphorical use. A re-presentation of a cat, let's say a picture of a cat, will not lead to a phenomenal experience of the cat. A re-presentation of a cat, inserted in some functional loop, will not lead to a phenomenal experience of the cat.

1.7 Phenomenal is intentional

We are tied to the environment. What is the purpose of having an inner world which is different from the external one? Many authors tried to develop an inner role for phenomenal experience as if phenomenal contents were severed from the physical world. Fostered by virtual reality and a Cartesian conceptual framework, an autonomous role for phenomenal experience had been defended in a wide spectrum of standpoints ranging from extreme qualiophily to more cautious dual properties stances.

2 Crucial issues

On the other hand, as to phenomenal consciousness there are several critical issues that are seldom taken into account and that ought to be critically examined. Phenomenal contents have several properties that do not match well with the physicalist framework. However, instead of using the mismatch to justify an easy

escape route into dualism or idealism, I will suggest that it is the sign of an incomplete and coarse understanding of the nature of the physical world. Consciousness could be the natural locus where to achieve a better understanding of the physical world. In such a way, phenomenal content could be the missing piece to complete the scientific jigsaw.

2.1 Unity is not easily explained in physical terms

More often than not, conscious experience is some kind of unity. I perceive myself as a self endowed with some kind of unity. My percepts appear to be other kinds of unities. I feel a unity between myself and the object of experience. However, the world does not seem made of unity. On the contrary, the reductionistic stance pulls towards the atomistic level. But, if everything has to be reduced to atoms, molecules, or atomic events, how can we explain the emergence of larger unities?

The same problem arises in biology. Living organisms seem to be somehow different from their parts. Is the cell a unity different from a heap of molecules? Many authors would energetically defend such claim (Varela, 1979; Maturana and Varela, 1987/1998; Thompson, 2007), yet no clear consensus has yet been reached. Others authors would prefer more reductionistic views. Isn't the cell just a collection of molecules?

Phenomenal experience presents the problem of unity in many different ways. Perhaps, the most evident is that of the unity of our phenomenal world which had been studied at length by Gestalt psychology. A more recent version is represented by the binding problem (Revonsuo, 1999). Neither of them had a solution yet.

2.2 Spatial localization

This issue was raised by Alfred North Whitehead (Whitehead, 1920; Whitehead, 1929/1978; Whitehead, 1933) and eventually echoed in several ontological conundrums. What is the spatial localization of a certain phenomenon? This is a simple problem when confronted with simple physical phenomenon such as the nuclear decay of a particle. But, what about perception or motor control? Where do they begin and where do they end? Shall we take the external object to be part of the process? Shall we take the peripheral nervous activity to be part of our mental processes or not? Even inside the brain, we could apply what has been called the problem of the many (Unger, 1980; Lewis, 1993): why should we take a group of neurons as part of the brain?

Moreover, if a phenomenon is spread in space, it means that it is made of separate parts. What glues together such parts (Nelson, 1995; Revonsuo, 1999)? What changes a group of scattered parts into a unified whole (Simons, 1987; Varzi, 1996; Atkinson, 1998)?

The same problem applies to potentially conscious machine. If we have several activations in the memory of a computer, why should they be taken as a whole? Is there any difference between the neural organization of a starfish and that of a mammal as to their being a unity? Recently, Giulio Tononi began to address this problem introducing the concept of integrated information (Tononi, 2004). However, more work is needed.

2.3 Temporal localization and length of present

The above problem can be applied to the case of the temporal location of a phenomenon. How can we define the beginning and the end of a phenomenon? This is not an easy problem since it is entangled with the notion of present. What is the present and how much time does it last?

Indeed, these are very deep waters. They hide an unresolved fight between neuroscience and physics. According to classic physics the present is instantaneous. Isaac Newton developed the calculus of first derivatives since he wanted to be able to express everything in a temporal present of null length. This is the most common notion of present: a Euclidean temporal point. This is way in system theory, scholars love to express the status of the system as a differential vectors at time t .

However this notion of present does not match either with neural phenomena or with phenomenal experience. Our phenomenal experience of time is thick as it has been famously and differently stressed by William James (specious present) and Henri Bergson (duration). However, it is often underestimate the fact that neuroscience too is incompatible with a present of null length. Neural activity takes time. Spikes take time too. If we could take a temporal slice of increasingly thinner thickness, all neural notions would disappear one by one. If a temporal activity spans many milliseconds, there must be some sense in which events distributed in that span of time belong to the same instant of time – i.e. to the same unity. This is often overlooked by many who rather uncritically assume the existence of temporal windows.

Many questions are left unanswered and, which is worse, unquestioned. Is there a limit to the useful length of the present? If we could slow down the brain activity, would we have a corresponding slowed down phenomenal experience? Would there be any limit to how much we could stretch the temporal scale? Is the present equally thick for all process or some process are faster and others are slower? Is there any causal criterion that defines the length of present for a given phenomenon? What is the relation between the phenomenal time and the physical time?

2.4 Presentation nor representation

Are we sure that during conscious experience, we represent the world? Couldn't it be that somehow the world present itself (Honderich, 2006; Manzotti, 2006; Albertazzi, 2007). Are we sure that there is a dichotomy between conscious experience and physical reality, between representation and represented?

2.5 Boundary of physical phenomena identical with phenomenal experience

I assume that phenomenal experience is a physical phenomenon. However what is the corresponding physical phenomenon? Neuroscientists automatically assumed that we should seek such physical foundations should among neural activities. However, there is no guarantee that such assumption is right. As a physicalist, I think that there must be a physical phenomenon identical to our experience. However, not being a neurocentric neuroscientist, I am a little more agnostic about where to look. Isn't it possible that a different physical phenomenon spanning more time and space would be our phenomenal experience (Manzotti, 2005)?

3 Conclusion

The just outlined overlapping and partially contradictory statements challenge the implicit background from which most of conscious research is carried on (in psychology, philosophy, and neuroscience). They simply outline the gist of current scientific research. This is precisely the point that Alfred North Whitehead observed (Whitehead, 1925: 48):

When you are criticising the philosophy of an epoch, do not chiefly direct your attention to those intellectual positions which its exponents felt it necessary explicitly to defend. There will be some fundamental assumptions which adherents of all the variant systems within the epoch unconsciously presuppose. Such assumptions appear so obvious that people do not know that they are assuming because no other way of putting things has ever occurred to them. With these assumptions a certain limited number of types of philosophic systems are possible, and this group of systems constitutes the philosophy of the epoch.

The problem of consciousness is going to jeopardize most of our basic assumptions about the nature of the physical world. In order to deal with such a troublesome aspect of our experience, we will probably need to revise them. In order to build a conscious machine we ought to understand what the physical foundations of phenomenal content are.

References

- Albertazzi, L. (2007), "At the Roots of Consciousness: Intentional Presentations" *Journal of Consciousness Studies* 14(1-2): 94-114.
- Aleksander, I. (2008), "Machine consciousness" *Scholarpedia* 3(2): 4162.
- Atkinson, A. P. (1998), Wholes and their parts in cognitive psychology: Systems, subsystems and persons., in *Proceedings of Wholes and their parts* Bolzano (Italy).
- Eddington, A. S. (1929/1935), *The nature of the physical world*, New York, MacMillan.
- Harnad, S. and P. Scherzer (2008), "First, Scale Up to the Robotic Turing Test, Then Worry About Feem" *Journal of Artificial Intelligence and Medicine* (Special Issue on Artificial Consciousness).
- Honderich, T. (2006), "Radical Externalism" *Journal of Consciousness Studies* 13(7-8): 3-13.
- Jammer, M. (1954/1993), *Concepts of Space: The History of Theories of Space in Physics*, Cambridge (Mass), Dover.
- Lewis, D. (1993), "Many, but almost one" in *Ontology, Causality and Mind: Essays in Honour of D.M. Armstrong*, J. Bacon, Ed., New York, Cambridge University Press.
- Manzotti, R. (2005), Outline of an Alternative View of Conscious Perception, in *Proceedings of Towards a Science of Consciousness 2005* (TSC2005)Copenhagen.
- Manzotti, R. (2006), "Consciousness and existence as a process" *Mind and Matter* 4(1): 7-43.

- Maturana, H. R. and F. J. Varela (1987/1998), *The tree of knowledge: the biological roots of human understanding*, Boston, Shambhala.
- Nelson, J. I. (1995), "Binding in the Visual System" in *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., Cambridge (Mass), MIT Press: 157-159.
- Revonsuo, A. (1999), "Binding and the phenomenal unity of consciousness" *Consciousness and Cognition* 8: 173-85.
- Simons, P. M. (1987), "*Integral Wholes" in Parts: A Study in Ontology*, P. M. Simons, Ed., Oxford, Clarendon Press: 324-360.
- Skrbina, D. (2005), *Panpsychism in the West*, Cambridge (Mass), MIT Press.
- Strawson, G. (2006), "Does physicalism entail panpsychism?" *Journal of Consciousness Studies* 13(10-11): 3-31.
- Thompson, E. (2007), *Mind in Life*, Cambridge (Mass), Belknap.
- Tononi, G. (2004), "An information integration theory of consciousness" *BMC Neuroscience* 5(42): 1-22.
- Unger, P. (1980), "The Problem of the Many" *Midwest Studies in Philosophy* 5: 411-467.
- Varela, F. (1979), *Principles of Biological Autonomy*, New York, North Holland.
- Varzi, A. (1996), "Parts, Wholes, and Part-Whole Relations: The Prospects of Mereotopology" *Data and Knowledge Engineering* 20: 259-286.
- Whitehead, A. N. (1920), *Concept of Nature*, Cambridge (Mass), Cambridge University Press.
- Whitehead, A. N. (1925), *Science and the modern world*, New York, Free Press.
- Whitehead, A. N. (1929/1978), *Process and Reality*, London, Free Press.
- Whitehead, A. N. (1933), *Adventures of ideas*, New York, Free Press.