

Some remarks about consciousness

Giuseppe Trautteur

trau@na.infn.it

Ordinario di Cibernetica, Dipartimento di Scienze fisiche, Universita' di Napoli Federico II

1. Taking stock of consciousness: not an inventory of theories, but an examination of the concept.

There simply are no viable theories of consciousness. (From now onwards: cs.) Many authors, and very distinguished at that, have put forward ideas about the workings of cs, but even the very notion of what could count as a theory of cs, or what is there to explain about cs is not clear. Even what is a fact of cs is marred by the first person/third person dilemma. The very notion of explanation with regard to cs differs from, for instance, the same problem with regard to physics. Objective results as, say, the perceptual delays which spurred the multithread “theory” of cs, or actual pathologies of cs as depicted all over Damasio’s *The Feeling of What Happens*, are scarce and far between, and therefore very important. They might well be the sole legitimate kind of studies on cs yet. Accordingly – and unfortunately – the present text is not.

2. Cs arises in, or inheres to, or is a property of some animal organisms.

Not knowing what cs might be – or even whether the verbal phrase ‘to be’ is applicable in that context – one has to be very careful in the use of predications usual with entities of assessed ontological status.

Also this is a disclaimer about possible claims that cs is indiscriminately present in many, most, possibly all, biological or perhaps even inanimate entities. Such animistic attitudes, while perhaps expression of positive pragmatic and ethical worldviews, are quite useless in advancing the knowledge of human cs.

However, it is entirely possible that artificial so-called information processing entities might become endowed with cs either deliberately or, more probably, incidentally in the course of the development of artifacts tending to simulate human behaviour.

3. The organism exchanges energy and matter with the environment. Different time scales: slowest = evolutionary (million to billion years), middle = development, adaptation, learning (minutes to years), fast = metabolism, current behavior (milliseconds to years). Different mechanisms are involved.

Comments to such a cartoon would require an entire course in biology and evolution. Its purpose is only to indicate the outlook which is being used in the present remarks on cs.

Notice only the opinion that groups together development, adaptation and learning and the use of time scales to characterize three conceptually and biochemically separable entities: the entire planetary biota, the individual organism (as defined by the

Some remarks about consciousness

appearance of membranes) tied with its environment and the organism's metabolism, both material/energetic and symbolic.

Memory, which pertains to fast, current behaviour is the tie with the much slower adaptive and learning behaviour.

Cs pertains only to the fast, current behaviour, although the contents of cs (in particular Damasio's autobiographical self and extended cs) need memory.

4. Quantum effects are present in their usual subsumption of macroscopic matter. However, there is no intimation of macroscopic coherent states to which a definite connection with cs could be ascribed.

Again, this is a disclaimer against claims that macroscopic quantum phase correlations of the wave function of some biological molecules are connected with cs. Another hypothesis, put forward many decades ago by no less an authority than John C. Eccles and still maintaining some currency, proposes as an explanation of freewill – closely, but not necessarily, connected with cs – the inherent non determinism of the wave packet collapse in the determination of the post synaptic potential at the synapse.

Contrairwise, phycisists are still exorcising the necessity of a conscious observer in order to obtain the said collapse.

All this sounds very strange, if not downright spooky.

5. Does the organism also exchange information?

Every schoolboy knows that this is so. And yet neither schoolboys nor cognitive scientists know what information is.

Apart from newspapermen, and common usage – where the meaning of 'information' is very close to news –, there are perhaps three areas where the word appears to be used technically in well defined ways: the communication engineers, the computer scientists and some varieties of biological scientists, in particular molecular biologists.

While the word is commonly used in loose talk in these communities, it turns out that when it is used technically it does not have any more its common sense connotation. Thus we have logarithmic quantity of information over probabilities in the Theory of Communication and Information (forget the allegedly fundamental identification with entropy), data structures in Computer Science and combinatorial schemes in Molecular Biology. Philosophical investigation of this problem only shows that any general definition of information is circular, unless one posits a conscious understanding subject as the root and origin of meaningful information.

The only sensible way out seems to be that of using 'information' at its common sense value and, when really tackling the problem of information exchange between organisms as, for instance, in perception studies, to realize that it is signals that are actually exchanged and that whatever else is carried by the signals resides, at the moment, in the eye of the beholder.

6. No real (meaningful) information outside of cs, but enormous development of Information Technology.

Giuseppe Trutteur

Therefore IT is actually data structure technology. Yet the intentional stance confers meaning to those data, and many suggestions on the nature of mental processes emerge from the advancement of the technology.

7. Uselessness of mind for cognition, which can be entirely mind-free. Or rather, subject-free.

When not many decades ago the term ‘cognition’ acquired undisputed dignity in Computer Science – mostly in AI – and in the psychological and philosophical communities, it was understood that it meant those mental processes that do not involve feelings, emotions, mood, and cs itself. In the positive direction ‘cognition’ was associated with deduction, theorem proving, planning, games, perhaps inference in general, memory, learning, pattern recognition and so on.

Common sense reasoning, as well as machine translation and speech-recognition – those great problems left unsolved by old AI –, constitute, perhaps, the dividing watershed between cognition as rationality and reason as true understanding. Here the use of words is still very strongly tied up with a classical outlook, where the distinctions between *vous*, *psyche*, *logos*, etc. addressed different problems from the ones we address nowadays. In particular the possibility of mental behaviour in non sentient artifacts remained strictly unthinkable, outside of the magical, until the eighteenth century. There was no need of separating the symbolic from the understood. Now we know better and we should, although we not always do, distinguish between performing a cognitive task and understanding what is going on.

And the huge result is that all “mental” activities on the cognitive side of the above watershed do not need, for their effective performance, the mind of a subject.

8. Completeness theorem as first and foremost result of AI?

If one takes AI as the endeavour of understanding the mental process through its *in vitro* or syntetic duplication, the span of logical research between Frege and Henkin, peaking in Gödel’s Completeness Theorem, can be seen as a major triumph of what in later decades will come to be called AI, insofar as the all important and undoubtedly mental activity of correct, i.e. truthful, logical deduction has been completely explained. After that came resolution, practical theorem provers, SOAR, etc., as well as psychological inquiries on how the human organism does, or does not, implement correct deduction.

9. First person and third person accounts of psychological facts is accepted.

A remark on the fact that S-R attitudes are a thing of the past and that eliminativists are just a sector, and a dwindling one at that, within the cs studies community.

10. The notion of phenomenal cs is generally accepted. Connections with the term ‘phenomenology’.

It looks as if the word ‘phenomenal’ acquired widespread currency in the recent cs studies community after Ned Block’s introduction of the access/phenomenal dichotomy of cs. In this sense the word refers to first person experience as mentioned

Some remarks about consciousness

above. However within the many accepted meanings of the word – for instance the physicists usage of ‘phenomenological theory’ to denote a theory which just fits the data with no explanatory connection with accepted first principles – there is the one which grounds the philosophical line of thought beginning which Brentano, established by Husserl and developing through, for instance, Merleau-Ponty and many others, to the present time. The cs studies community has renewed the interest of phenomenology in cs mainly through Brentano’s intentionality.

11. No psychophysics of the intentional kind: if one accepts intentionality, the game is changed. Besides, the Schoolmen (or Lonergan) already had a (dualist) very satisfactory psychology. Vestige of animistic thought? Folk psychology is deeply dualistic. We all are.

As is well known, at its origins in the second half of the nineteenth century Psychology practiced introspection together with very sophisticated experimental procedures (von Helmholtz, but particularly Wundt, Fechner, etc. marvelously capped by James), establishing the basis of perceptual psychology. The term Psychophysics was then introduced and with it the idea of the parallelism between the physical (human) organism and its mental experience. If intentionality, or aboutness, is the mental ingredient of phenomenological analysis, then – whatever denial may be opposed to it – some sort of dualism is on the horizon. Perhaps not so-called substance dualism, but certainly at least the wafting of something immaterial.

If we accept the naturalization of cs, then an essentially experiential and first person element as intentionality does not count as an explanation: it begs the question.

Still, the origin of any mental activity resides certainly in first person experience. This is also the unavoidable everyday stance (i.e. the intentional stance which, in a reciprocating gesture, we assume towards our fellow humans, some animals and some artifacts).

And yet one insists in explaining this first person experience in the usual third person way, because it works so well in every scientific context but this one.

Then one must be prepared to live within a possibly vicious circle and in lingering contradiction until the discovery of something radically new.

12. Complexity relates to cs in an ancillary way, but does not determine it. Neither self-organization, whose theoretical possibility is a consequence of recursion theorems, but is not a “level” of matter organization.

What we know in general about complexity is that it has to do with being large, according to some dimension, as well as intricate – we border on the tautological. Actual technical uses of the word, mostly in computability theory and in Santa Fe (the two communities scarcely even know of their reciprocal existence), capture very limited, although sometimes stupendous (e.g. Julia sets), aspects of the every day connotation of the term. Hoping that complexity will give raise *per se* to cs looks as if shaking a test tube of some alchemic concoction will generate a homunculus.

Yet nobody has any doubt that a conscious system is going to be rather large, perhaps enormously so, and very intricate. The CNS of conscious organisms certainly is such.

Giuseppe Trutteur

The level metaphor comes from conservative fields, in particular gravity, and from our proprioceptive experience in resisting it. It is a useful, but dangerous, metaphor because, without further specifications, it does not add anything explanatory at what is being stratified in levels and may lead into unwarranted substantializations.

13. Probably cs is best studied in relation to the current, fast behavior of a semi-adiabatic physical structure.

Probably cs is based in physical systems of an essentially classical, macroscopic, deterministic nature, with probabilities arising only for thermodynamical reasons.

Adiabatic is used in its mechanical, not thermodynamical, sense, and refers to certain stability properties in the behaviour of mechanical systems. Orbits, in particular. These stable orbits acquire, accordingly, a certain substantiality and independence from the actual entities whose motion defines them. They may be used as building blocks in the structure of further entities. Here the level metaphor almost becomes literal.

This idea is tied in with the notion of virtuality introduced below in n° 20.

14. The traditional hierarchy: awareness, “plain” cs, higher order cs.

It turns out that the locus of qualia, experience, subjectivity, etc. is “plain” cs. The elucidation of “plain” cs is the problem correctly identified by Chalmers as hard. “Higher order cs” is usually associated with introspection, rationality, extended memory, etc. and is considered as less common and more valuable. Hence the aristotelian ordering. But most of the prerogatives of higher order cs belong to Chalmers easy problem. Still the ordering is unswervingly adhered to in the literature.

15. In English the word ‘awareness’ muddles the problem. In Italian ‘stato di veglia’ is less committal.

Again, it is very hard to distinguish awareness from cs proper. Non human animals might or, again, might not be conscious. But it seems correct to consider them aware in the sense that they are alert and reactive. Unfortunately the awake state (stato di veglia) brings to the fore the problems of the nature of sleep and of the status of cs during sleep, dream and, if they exist, the altered states. Problems perhaps larger than that of cs and which are not being considered in this text.

16. Awareness, attention, and absentmindedness.

It is safe and economical to assume that there exists just one cs, i.e. “plain” cs, and that self-cs is not cs of a higher level, nor cs of the self – whatever the exact meaning of this form of words might be –, but just cs, and that awareness is synonymous with cs, and that when we want to refer to a form of behaviour that is not quite conscious, but has a reactive and alert quality to it, we just use reactive behaviour.

The relationship between attention and cs is complex. There may be attention without cs in the sense that a specific task, selected between a, usually large, number of them, is being correctly performed (attended to). The notion may be used even for robotic artifacts. And it may be that the organism consciously concentrates, focuses,

Some remarks about consciousness

makes salient some particular aspect or fragment of its (his) field of perception or action. And, perhaps, ignores others concurrent parts of its behaviour. The usual example is that of humoring a conversation while driving: where is the attention directed to? And an actually absent-minded (sic!) organism may suddenly be drawn back to attention by some special sensorial stimulus (or internal process): as in system interrupts or polling in computer programming: absent-mindedness, obviously, can never be total, except perhaps in extasis or serious psychotic conditions.

These very important questions do not really touch the problem of cs. Their study will certainly bring important factual elements to the understanding of cs, but in fact all the above situations concern the problem of conscious behaviour which already presupposes the presence of cs in the organism.

17. Damasio's hierarchies: i) basic life regulation / emotions // feelings / high reason; ii) proto self // core self / autobiographical self; iii)

Damasio's core cs is "plain" cs and, according to him, takes place between emotions and feelings as well as between proto self and core self (marked by double slash). Extended cs arises between feelings and high reason as well as between core self and autobiographical self.

Damasio follows the conventional ordering, but by stressing the chasm between emotion and feeling, clearly considers "plain" cs the fundamental issue.

18. Introspection vs reflexion. Introspection as metalanguage and self-reference is well understood and machine instantiated.

Introspection presupposes an agent, possibly conscious, that analyzes itself. Problems arise when analyzing those parts or processes that are active in the analytic process itself. These problems are akin to undecidability results in computability theory and incompleteness results in logic. They are well understood and give rise to very important applications in operating systems and complex software systems. No cs (nor self-cs) is involved.

Many authors of neurological upbringing maintain that cs is tied with such self-referential and metalinguistic activities via higher order thoughts. A HOT is a thought about another thought. If thoughts are seen as just symbolic activities, then machines have HOTs and should be conscious. Some authors accept tentatively this consequence, but insist on some kind of Harnad-grounding of the symbols (E.K. Rolls); others (e.g. Damasio in *Looking for Spinoza*, pp. 126-130), although overtly naturalizers, maintain that this is cannot be so because the machine and its components are not alive, they lack some biological quality.

On the other hand, reflexion is taken to denote that total folding over ourselves that we experience as our own identity.

Disclaimer: Reflexion seems to be a necessary condition for cs. We experience ourselves reflexively as a condition of possibility for experiencing in general. But reflexion is an introspective requirement, not an assessed reality. And even if it were physically understood and perhaps artificially implemented, it would not, *per se*, entirely explain cs. The major question of why and how there are feelings, and an

Giuseppe Trutteur

owner of them, would remain open. Thus it is not a theory of cs. Trying to delve with reflexion seems, however, to inflict a little dent to the hard problem.

19. Difficulties with reflexion: first attempt: it is the presence o the self to itself; second attempt: the identity of the owner (of experience) and the owned; third attempt: if conscious experience is a process, then that process is experiencing itself. This is how the process constitutes the subject.

All these three “attempts” border on a violation of the identity principle.

Assuming that the CNS is a symbol processing device (*pace* of supporters of the view that it is a non symbolic dynamical system and that neural networks are not algorithmic), there is a path of research further exploring this algorithmic nature, for which matter does not appear to be quite so strict a principle of individuation.

Damasio, *The Feeling of What Happens*, p. 170 seems to be the only reference in the literature that actually tackles this reflexive phenomenon. The gist of his argument is that the operation of the brain “represents its own changing state as it goes about representing something else”. So there seem to be two sorts of contents of cs: the “representing something else”, which is the usual content, and the “represent[ation(s) of] its own changing state”, which is cs itself. All this is rather metaphorical, as acknowledged by Damasio himself.

20. Levels of virtuality, material grounding. When is it correct to speak of levels? The Universal Turing Machine theorem is crucial, but the relevant issue is the effectivity of codes: running a program is indistiguishable from the action of an actual machine. Interpreters

Again a cartoon. This time about computability theory. The importance of the Universal Turing Machine, in this context, is not its universality – i. e. that it behaves as *any* other TM, given the code –, but that there *exist* codes (i.e. symbolic, effective - i.e. material -, entities variously called gödel numberings, programs, software, implemented algorithms, etc.) that allows a single physical machine to effectively behave as another one.

In particular as another UTM.

The stratification of levels of these successive UTMs, starting form a physical one, for instance a Pentium (and this is the actual grounding of symbols perhaps akin to, but different, from the one introduced by Harnad), is a legitimate hierarchy of levels, albeit a tangled one, as observed many years ago by D.R. Hofstadter.

The UTM of the successive levels as well as all the machines not actually present, but *interpreted* by proxy by some UTM, are called virtual machines.

This is the meaning of ‘virtual’ in the present text.

21. Simulation is not the same as virtuality. The former is analogy and sameness *secundum quid*. The latter tends to be equality without residue; full substitutivity..

The ontological status of a running program as distinguished from the activity of a physical machine is a huge open metaphysical problem. Which category captures the mode of being of a running algorithm?

Some remarks about consciousness

- 22. Identity between two things which are the same thing which is involved in playing different roles at the same time and in two contiguous levels of virtuality. Physical instantiation of virtuality in artificial systems and in biological neural tissue. Possible relevance of fix-point theorems of computability theory.**

Vague suggestions of the possibility of reflexion. Work in progress.

- 23. Cs as a process (James) is banally understandable today because through virtuality we are becoming used at looking at a process as upon a substantial entity. Before that, the notion was rather metaphorical.**

But even substance – that which is below, concealed, immutable – is not a basis anymore: yes there is the biological CNS and the electronic CPU. But resorting to them does not count as an explanation. It was called reductionism. What are we now going to hang on?