

Consciousness is localized and we can discover where it is

interview with Christof Koch* by Andrea Lavazza**

*Christof Koch, 44 years old, is professor of Computation and Neural Systems at the California Institute of Technology. He has just published *The Quest for Consciousness. A Neurobiological Approach* (Roberts&Company Publishers, 2004) in which he summarizes years of empirical research, started with Nobel laureate Francis Crick, on the molecular and neural basis of consciousness.

**a.lavazza@avvenire.it

You adopt a vague definition of consciousness as scientific object to investigate. Could you explain what is consciousness to you and why you have chosen that definition?

"It is a capital mistake in the early phase of the scientific investigation of a phenomenon to try to define it in a too narrow fashion. Philosophers for the past two-and-a-half millenia have attempted to define consciousness without much success. If you can read these lines, you are conscious. In an operationalized sense, at the very least, a conscious subject - whether a person or an animal - must be able to selectively attend to events in the world and to store behaviorally relevant information for a brief amount of time. Throughout my book, I use awareness and consciousness as synonyms. According to John Searle, consciousness consists of those state of sentience which begin in the morning when we awake and continue throughout the day until we fall into coma or die or fall asleep again".

What is the basic assumption (hypothesis) that guides your research on consciousness?

"That we have the tools and techniques at our disposal to search for the minimal neuronal mechanisms jointly sufficient for anyone specific conscious experience, the neuronal correlates of consciousness (NCC). While it remains an open question whether discovering and characterizing the NCC will be sufficient to understand the structure, function, and origin of consciousness, it is a necessary step".

You and Francis Crick have produced a framework for consciousness that in your book you enrich with new data. Could you summarize it in its main aspects?

"These are summarized in our 2003 article in *Nature Neuroscience*, 'A Framework for Consciousness' (vol 6, no2). First, we take consciousness seriously, as a brute fact that needs to be explained. The first-person perspective, feelings, qualia, awareness, phenomenal experiences - call it what you want - are real phenomena that arise out of certain privileged brain processes.

We speculate that the actual NCC may be expressed by only a small set of neurons (competing cellular assemblies), in particular those that project from the back of cortex that are not purely motor and that receive feedback from there. However, there is

Consciousness is localized and we can discover where it is

much neural activity leading up to and supporting the NCC, so it is important to study this as well as the NCC proper".

The first goal of your quest is to find the neuronal correlates of consciousness (NCC). What are they? What stage is the research at nowadays?

"In the case of visual conscious perception, it is known that neurons in the retina are not part of the NCC. That is, you don't see with your eyes but with your brain proper. While neurons in the retina are necessary for normal forms of seeing (but not for imagination or for experiencing visual dreams) their firing activity does not correspond to the way we see. This is widely accepted. What is more controversial is the proposal that Francis Crick and I made a decade ago, that the neuronal activity in primary visual cortex - the terminus of the visual information streaming up from the retina, and the first cortical region dedicated to vision - is not represented in visual consciousness. That neurons in V1 are not part of the NCC. However, in the meantime much evidence has accumulated in favor of this hypothesis.

This is worthwhile for two reasons. Firstly, it demonstrates that not any neuronal activity in the cerebral cortex contributes to consciousness. Only certain, privileged neural activity does. This argues against holistic ideas about consciousness. That consciousness emerges from the entire brain, that it can't be intimately associated with only a subset of the brain. Secondly, our proposal demonstrates that true progress in these difficult questions can be made.

The research program of many brain scientists throughout the world focuses on various aspects of the NCC. Which neurons or which brain areas are specifically involved in representing what aspects of phenomenal consciousness. The search for the NCC is tractable, clearly defined, and will yield to a concerted scientific attack".

Will focusing on visual sensation pave the way to explain the whole problem of consciousness? And why?

"It is likely that all the different aspects of consciousness, sensory perception, introspection, the emotions, self-consciousness and so on, all share one or more basic aspects with each other. And that once we understand the physical basis of one of these we will have made a leap in understanding many of these as well. That is, we believe that once we understand how the brain can be sufficient for specific conscious visual state the explanation for how the brain can generate the feeling of 'being me' or 'being angry' can't be far ahead. Working with visual perception has a number of advantages. Most important are the following two.

Psychologists have perfected a variety of techniques and illusions that allow them to manipulate the relationship between a physical stimulus and its associated conscious, visual perception. In some, you are looking at something but don't see it (motion-induced blindness; change-blindness; inattentional blindness). Or you may look at an unchanging input yet sometimes you see one percept and sometimes another one (bistable percepts as in a Necker Cube, flash suppression, binocular rivalry). This allows us to track the footprints of consciousness in the brain.

The visual system of monkeys is very similar to that of humans. Furthermore, monkeys can be trained to perform in many of the same experiments that visual psy-

Cristof Koch

chologists ask people to do. This permits the intrepid neuroscientists to explore the neural basis of visual perception in an invasive manner not possible in people.

Do you believe that a conscious-ometer could be designed and used to detect who and what is consciously aware? How long will it take? And, in your opinion, shall we have an experimentum crucis (an operationalized test) which can solve the problem of consciousness?

"Yes. Such a conscious-ometer will enable medical personnel to monitor the presence of consciousness in premature babies and young infants, in patients whose minds are afflicted with severe autism, or senile dementia, and in patients who are too injured to speak or even to signal. It will permit anesthesiologists to better practice their craft. Understanding the brain basis of consciousness will allow scientists to determine which species are sentient. Do all primates experience the sights and sounds of the world? All mammals? All multi-cellular organisms? This discovery should profoundly affect the animal rights debate since species without NCC can be thought of as bundles of stereotyped sensory-motor loops, without subjective experience, zombies. Such organisms could be accorded less protection than animals that do show NCCs under some conditions. I don't know how long this will take. Probably a few decades.

I don't know of any single critical experiment in any science that proves or disproves an entire theory. It is the cumulative force of many experiments that convinces".

What are at the moment the most important empirical facts on which you rely when you talk about of consciousness in terms of NCC?

The most important empirical facts in the evolving debate on the neuronal basis of consciousness comes from listening in on the electrical activity of individual nerve cells in monkeys that are trained to carry out simple perceptual experiments. Neurons are the relevant microvariables that underly perception, thought and action and we need to understand how they respond. Another source of new data, but at a much coarser spatial and temporal resolution, are experiments with human subjects placed inside a magnetic scanner. This allows us to observe their brain activity while they perceive or remember things. A third source of information are detailed clinical investigations of specific neurological deficits in patients. I wrote a 300 page book summarizing all those salient facts that are of relevance to understand the neural basis of consciousness".

One can say that the problem of the treshhold and the problem of detecting the few key neurons sufficient for a particular feeling are the unknowns of your framework. Do you agree?

"The conscious perception of anyone stimulus attribute, say the gender of a face or the bark of a dog, appears to be all-or-none. That is, you either see its gender or you don't (of course, any percept in the real world has many different attribute and each one is all-or-none). This could most easily be explained by postulating that the neuronal coalition sufficient for this percept has to established itself for a minimum

Consciousness is localized and we can discover where it is

amount of time. And that this requires that their firing activity exceeds some threshold. Where are these neurons that make up this coalition? Do they inhabit a particular brain region? Is their one or more commonalities among the neurons that are sufficient for the percept of 'red' and 'smelling a rose'? Will they all be of the same type? Do they project to the same brain area? Do they exist in new-born babies and so on".

Why and how should the short term memory have a key role in explaining consciousness?

"In both healthy people and patients, the existence of some level of consciousness goes hand-in-hand with the ability to store some non-stereotyped information for a brief while, say tens of seconds. Thus, the requirement for some sort of short-term buffer of information".

Someone says that your framework is too concerned with information flows in the brain and not concerned at all with the inner experience as we feel it. As a critic of yours put it, the question is: why should there ever be anyone at home inside such a brain?

"Francis and I have put aside the question at the heart of the mind-body problem - why does phenomenal experience feel like anything? We do that for now since speculations over the centuries have not led to any resolution of this question. Once the NCC is understood, this problem needs to be addressed again. In the meantime, we can solve many of the related questions associated with the mind-body problem".

You said that you and Crick used ideas that have been suggested before (i.e. Dennett's, Baars', Edelman's...), but the combination is original. Someone doesn't agree. What do you reply?

"Many of our individual ideas have been discussed over the past century by other scientists. What is relatively unique is our emphasis on the crucial importance of local properties of synapses, neurons or brain areas. And, of course, our specific hypothesis such as the 'VI' hypothesis discussed above. Nobody has proposed something like it".

What is your reaction to these labels referred to your theory of consciousness: monism, materialism, emergentism, epiphenomenalism?

"I think most of these philosophical labels are silly and too simplistic. I am not beholden to any one particular ideology. I study the existing natural world using the scientific method".

Once the NCC are discovered, will it be possible to understand how consciousness arises within our brain? And then, shall we have the know-how to create an artificial consciousness?

"Discovering the NCC would constitute a major step forward on the road to a final understanding of consciousness. Identification of the NCC would enable neuroscientists to manipulate its cellular substrate on the basis of pharmacological intervention

Cristof Koch

and genetic manipulation. It may be possible to fashion transgenic mice whose NCC can be rapidly and safely switched on and off. What behaviors might such zombie rodents be capable of? Clinical benefits will flow from this discovery as well, such as a better understanding of mental diseases and the design of new and powerful anesthetics, with few side effects.

Eventually, a theory that bridges the explanatory gap, that clarifies why activity in a subset of neurons is the basis of (or, perhaps, is identical to some particular feeling, is required. This theory needs to make comprehensible why that activity means something for the organism (e.g., why does it hurt?) and why qualia feel the way they do (e.g., why does red look one way, quite different from blue?). Such a theory will also predict which artificial systems, say the Internet as a whole or some autonomous vehicles, could experience something, could have conscious states (or why this might not be possible)".

What is today the single empirical puzzle you would like to solve in order to take a big step in your quest for consciousness?

"Within the cerebral cortex, there are massive axonal connections, the wires that permit the electrical pulses generated by nerve cells to travel from one neuron to the next. This information streams not only from lower cortical regions into higher ones but also back. I believe that if all such feedback connections are temporarily switched off, inactivated, that conscious states would not be possible anymore. That the subject might still show meaningful behavior, but only to learned, stereotyped situation and would not experience anything (similar to a zombie). Given the very rapid progress in molecular biology, it is possible that we can test such a proposal within the decade in mice or monkeys".

What is the relationship between consciousness and mind? Does the problem of causal efficacy on brain of the internal states of mind persist in your framework? Or does it fade away as mind and brain are the same thing?

"At this point we don't understand the exact relationship between the mind and the brain. It is clear that without brain, there is no mind and that any change in a mental state is associated with a change in a brain state. So there is no question that the two are heavily intertwined. Whether the two are the two sides of the same coin, as some assert is not clear at this point. The character of brain states and the character of phenomenal states appear too different to be completely reducible to each other. I suspect that their relationship is more complex than traditionally envisioned".

Does your theory of consciousness affect the great old topics of reference and meaning as they are investigated in philosophy of language? And what can you say regarding the aboutness, often considered the distinguishing feature of the mind?

"The aboutness of different states of consciousness, their meaning, derives from the penumbra of the neurons that are sufficient for anyone conscious state (The "penumbra" or "fringe" of consciousness, in William James' language, is all those conscious events that are not like coffee cups and musical melodies. They are "feelings of

Consciousness is localized and we can discover where it is

knowing," esthetic judgments, senses of right and wrong, all experiences that are hard to characterize, but which people can often report with high confidence and accuracy, *editor's note*).

The objects I feel, see, or hear out there in the world are not meaningless symbols but come with rich associations. The bluish tinge of a porcelain cup brings back childhood memories. I know I can grab the cup and pour tea into it. If it falls to the ground it will shatter. These associations don't have to be made explicit. They are built up from countless sensory-motor interactions with the world over a lifetime of experiences. This elusive meaning corresponds to the sum total of all synaptic interactions of the neurons representing the porcelain cup with neurons expressing other concepts. All the vast information are symbolized, in a shorthand way, by the qualia associated with the percept of the cup.

The penumbra is a term that stands for the neuronal processes that receive synaptic input from the NCC, without being themselves part of it. The penumbra includes the neural substrate of past associations, the expected consequences, and the cognitive background of the conscious percept. The penumbra provides the meaning, the aboutness of the percept (synaptic effects and firing rates). Qualia come to symbolize all of this vast, explicit or implicit, information contained in the penumbra".