PSYC515

Term-Paper

Approaches to the Study of Consciousness

Daniel Nikovski

Southern Illinois University at Carbondale

December 10, 1994

Running title: Approaches to the Study of Consciousness

Abstract

The paper explores the ramifications of the theory of consciousness and presents recent views and approaches towards its explanation. Two definitions of consciousness are distinguished - a more general definition commonly used in philosophy and a more rigid psychological/neurophysiological definition. It is argued that consciousness does not exist as defined in philosophy - to support this view, analogies from the history of science are brought forward. Instead it is suggested that the term consciousness needs to be replaced with a more precise psychological term such as "self-awareness" or "self-perception" which can be studied with methods from the natural sciences rather than seeking explanations in metaphysics. This approach is supported with recent developments in the psychology of consciousness and neuroscience. Several alternative approaches are reviewed and analyzed as well.

Defining Consciousness

It is most natural to start with a clear definition of the scope and meaning of what should be studied in this paper. This, however, is surprisingly difficult when it comes to defining precisely the meaning of the concept "consciousness". The main reason is probably that many scientific and non-scientific disciplines make frequent use of this term, ascribing different and often contradictory meanings to it. This paper will exclude from consideration the interpretations from non-scientific disciplines such as law, religion and ethics. However, even when scientific stance is adopted, at least two different interpretations remain.

Philosophy treats consciousness as a quality that distinguishes thinking from non-thinking matter. To philosophy of mind this distinction is of central importance and the term "consciousness" is mainly used to this end (Churchland 1984). Since philosophers try to adopt as wide perspective on the studied phenomena as possible, their definition of consciousness is of necessity too broad. All mental phenomena fall into the category "consciousness" - even processes that have been classified and studied individually in psychology like perception, attention, categorization etc. Accordingly, the question "How is consciousness possible?" in philosophy should be understood as "What makes it possible that certain objects can think while others cannot?". In this meaning "conscious" is simply a synomim to "mental".

A much narrower meaning is ascribed to consciousness in psychology and neuroscience. William James equated consciousness with attention and due to this tradition attentive processing and conscious processing are treated as synonims (James 1883). Consciousness refers then to a specific mental resource that is allocated to other tasks and processes that cannot be automated. This resource assists their automatization, presumably by enhancing the necessary associations. As such it is not a synonim to neither learning, nor skill acquisition, nor remembering; rather it is a higher-level process that schedules resources between the lower-level ones.

Farthing (1992) further subdivided this narrower meaning of "consciousness" into three possible interpretations - awareness, wakefulness and an executive control system, and tried to limit the use of "consciousness" to awareness only. However, these need not be necessarily considered as distinct interpretations. Rather, awareness and wakefulness are the result of the abstract mental process described above, while central executive system is the function (i.e. job) of this process.

Given this perspective, it can be seen why consciousness can be interpreted as self-awareness. Self-awareness arises when the subject and object of cognition coincide. In particular, we are conscious when we can "observe" as with an internal eye the processing taking place in our minds. For example, we are conscious when we know that we are doing something, that is, we are aware of our own actvities. We can hypothesize that many low-level cognitive activities are taking place at any specific moment whose task is to know the outer world, while consciousness is a cognitive process of similar nature whose task, however, is to know the operation of the other processes.

Narrowing in this way the scope and subject matter of the study of consciousness, the next step is to find a method that possesses enough scientific rigor to yield results. A temptation to be resisted is to plunge into purely psychological and metaphysical cerebrations. It was mostly due to the vagueness and non-scientific method of introspectionist psychology that the study of consciousness was neglected during the era of behaviorism. For this reason, a description at a purely psychological level is not likely to bear fruit.

Similar caution should be taken with approaches from the other end of the spectrum of psychological methods - the one occupied by behaviorist theories. Behaviorism is by definition incompatible with the notion of mental events, thus with the study of consciousness too. What is needed is an approach that will apply the scientific rigor of behaviorism to the mental phenomena that cognitive psychology has identified and tried to study so far.

The major hope in this regard comes from neural science. Currently a massive movement is witnessed that bridges the gaps between neural science, cognitive psychology and computer science. Different psychological phenomena have been approached quantitatively from neurophysiological point of view. These approaches typically boil down to implicating certain neural structures in mental phenomena and simulating their operation computationally. The next section discusses one such approach proposed by Koch and Crick.

The rest of the paper is organized in order of moving from neuroscience through connectionism and cognitive psychology to philosophy, i.e. in order of increasing abstraction of the operation of mental processes.

Awareness Explained as a Binding Mechanism

The neurophysiological perspective on consciousness implicates concrete neural processes as substrates of conscious activity. Koch and Crick (1990) proposed that awareness is closely related to the **binding mechanism** in brain activity. The binding mechanism is responsible for creating a mental entity by combining information processed by different regions of the cortex. For example, in the visual system color and shape are processed along different pathways - the binding mechanism has to put the two types of data together somehow. Crick and Koch argued that the binding mechanism is the aforementioned high-level resource that cannot be allocated to more than few sites due to constraints arising from the construction of the brain. Binding is still a hypothetical process and no concrete mechanism that implements it has been demonstrated yet. Nonetheless, particular 40Hz oscillations in the cortex were implicated as the means of the binding mechanism. Even though the actual mechanism will probably turn out to be much more involved, Crick and Koch's idea illustrates the approach that is taken in neuropsychology to explain consciousness.

The discussion above might as well serve as an explanation of attention rather than consciousness (which of course is not a problem if we assume that William James is right that consciousness *is* attention). We can hypothesize that consciousness is a special form of attention that arises when the object of observation is the self rather than the outer world.

Crick and Koch proposed concrete paths to study consciousness. They argued that visual awareness is a good subproblem to start with. The rationale behind this idea is the fact that the visual system has been already well studied and mapped in both humans and animals. If eventually the neural mechanisms that underlie visual awareness can be elucidated, this might give insight into cognitive phenomena of higher level like self-awareness that would otherwise be more difficult to study directly.

Crick and Koch made two basic assumptions. The first is that at any one moment some active neuronal processes correlate with consciousness, while others do not. That is, they assumed that consciousness is something that is related at different times to different more basic neuronal processes. It is also probable that some neuronal processes are never conscious - for example procedural memory (the memory that is responsible for movement and other highly automated skills).

The second assumption is that the different aspects of consciousness like sensory awareness and introspection share the same mechanism. This assumption is very important. If it is true, we can consider attention and self-awareness to be manifestations of the same process, but with different objects - the outer world for the former and the cognizing agent itself for the latter. If one of these manifestations is understood and explained, we can hope that the explanation can be applied to the others too.

Crick and Koch acknowledged the importance of modeling consciousness at the cognitive level, but considered the influence of symbolic models inspired by the von Neumann architecture of digital computers outright "pernicious". The dissimilarity between the way symbolic computers transmit and process information and the neurobiological evidence about the type of code used in real neural networks suggests that the operation of the brain can hardly be modeled with boxes labeled "CPU", "character buffer", etc, that are omnipresent in cognitive modeling.

Cognitive psychology can offer some very relevant models, though, that agree to a large extent with the view of Crick and Koch. Johnson-Laird (1983; 1988) suggested that the brain is a hierarchy of parallel processors and there is an operating system at the top of the hierarchy. The work of the operating system corresponds to conscious processing. Consciousness need not enjoy a special status - it might be just one of these processes, which works on other processes instead of on the outer world - a cognizer of cognizers, in a way.

The best point about this last view is that it explains very well why consciousness can provide *some* general information about certain cognitive activity, but not the *details* of this activity. Consciousness is simply a supervisor of the lower process that can observe the input and output of the process, but not its inner workings. Thus the conscious mind receives abstracted information from different cognitive processes.

This idea might be used to answer the question if consciousness can explore itself. If consciousness is a process itself, then it can be the object of monitoring too. But if only abstracted information is available to the observing process, consciousness cannot explore itself in enough detail. Indeed, if I try to think about my consciousness right now, all I can perceive is that this is what I am doing - thinking about my consciousness. I have no idea about what exactly my observation consists of. Thus knowing the inner workings of the self by introspection is probably in principle impossible.

The view that consciousness is a process that abstracts information from low-level processes is supported by Crick and Koch:

> We have suggested that one of the functions of consciousness is to present the result of various underlying computations and that it involves an attentional mechanism that temporarily binds the relevant neurons together by synchronizing their spikes in 40Hz oscillations.

The first part of this position, saying that consciousness abstracts the result of the work of low level processes has enjoyed much better appreciation than the second part that implicates 40Hz oscillations in the binding mechanism. Walter Freeman (1975) had proposed an alternative explanation based on nonlinear dynamics and chaos theory. To him the rapid change of what the thinking subject is aware of is due to the chaotic character of the operation of the brain. As it is known from nonlinear dynamics, small variations in the input of a chaotic system can change abruptly its equilibrium state, so in principle this is a possible mechanism too.

It can be noted also that consciousness, just like attention, cannot be concentrated on more than one thing at the same time. For example, I can either know that I am writing something at this moment or know what I am writing, but not both simultaneously. I can switch quickly enough between the two states, but this will not bring me to a state when I am conscious of the two things at the same time. This reasoning supports the view that consciousness *is* attention, but directed inwards. This can be further supported by the fact that consciousness seems to be a serial process just like attention.

Neural Darwinism

It is interesting to know not only how consciousness works, but also how it emerged, i.e., what processes took place to produce a conscious entity from nonliving matter. In this regard a good starting point might be to explore the fundamental mechanism that organizes individual neurons in neural networks. Of course, to adopt this view we must assume that it is the highly specialized processing that these neural networks perform that gives rise to mental phenomena. Gerald Edelman proposed a theory that gives account for this mechanism. He named his theory Neural Darwinism to reflect the fact that he was influenced strongly by the theory of evolution (Edelman 1987).

This theory contends that specialized mental activity is result of competition between groups of neurons to perform this activity. Different groups of neurons possess different **repertoire**, i.e. reaction to various signals. The group that is most fit to perform a function becomes specialized in it. Subsequent refinements are possible by using the same mechanism on progressively larger structures.

A direct consequence is the emergence of a layered architecture, the upper layers of which correspond to high-level processes, including consciousness. On a larger scale, Edelman modeled re-entrant circuits that give rise to cognitive phenomena like discriminations, memory structures, coordinating sequential steps in problem solving, etc.

Edelman's work should be noted for the attempt to bring together explanations on neurophysiological, cognitive and philosophical level, unlike most of the other approaches to the study of consciousness.

Connectionism - the Middle Ground

Connectionist models stand one step above neurophysiological theories in their level of abstraction of mental processes. Instead of studying real neurons, these models employ abstract units that can have the degree of similarity to real neurons that the studied phenomena require. Unlike cognitive psychology and AI models, connectionist models can make use of nonlinear dynamics and learning in a very natural way and at the same time discard those biological details that are not immediately relevant to the function that the neural circuit perform.

Not many connectionist models of consciousness have been proposed so far, which can be explained namely with the middle position of connectionism. While it is completely admissible in cognitive psychology to speak of vague and abstract modules like Central Executive, Phonological Loop, etc., the use of such "miracle" units is not allowed in connectionism. Instead, every neural network model has to be constuctively evaluated, i.e., its operation should be simulated on a computer and analyzed. As can be expected, connectionism so far has made much faster progress in explaining the lower level of mental processing - perception and motor control - than in providing insight into high-level phenomena like consciousness.

At the same time, connectionism cannot use experimental and clinical data like neural science. While neuroscientists can identify a mysterious process in undoubtedly conscious subjects (like the 40Hz oscillationsi in human cortex) and implicate it in consciousness, this approach is not possible in connectionist models again because of their constructive nature - we do not have yet a conscious neural network whose operation we can analyze.

It can be expected, though, that this middle ground is the right level of detail for the study of all mental phenomena, the place where "miraculous" models from cognitive psychology and AI can meet concrete neural mechanisms.

Baars (1994) pointed out that PDP modelers have provided domain-specific models (for perception, motor control, etc.) that can be used in higher-level modeling as building blocks. Baars considered the blackboard model of reading proposed by McClelland (1986) as an example of a connectionist implementation of the Global Workspace system (discussed below).

There have been criticisms to the connectionist (computational) approach to the study of consciousness (Revonsuo 1994; Searle 1990). However, these objections are on ontological basis and are not very convincing. For example, Revonsuo claimed that "computationalism is an inadequate ontological basis for theories of consciousness because ... computational systems cannot have different levels of organization in the sense that physical systems can and do." This claim is far from being true, to say the least. Much to the contrary, all computational systems currently used are deliberately designed to have levels of organization and abstraction. Revonsuo also argued that computation does not exist on its own - rather, an interpreter is necessary so that he or she can affirm that a computation is taking place. In this conclusion he was right, but he used it to claim that this is the reason for which no computation can possibly take place in the brains of thinking creatures. This last conclusion, however, is wrong. Of course an observer is necessary to analyze what computation is going on - this is actually what all scientists in the quantitative disciplines do - formulate quantitative models. An astronomer that studies the revolution of the Earth around the Sun would never say that the Earth computes anything, but a nevertheless will make use of computational model to predict its motion. Why would it be impossible to construct a computational model that could predict the results from the operation of mental processes?

Searle, as usual, has an objection that is not really an objection: "Computational models are useful in studying the brain, but nobody believes that the computational model is a substitute for a scientific account of how the system actually works, and they should stop pretending otherwise.“ Well, a computational model *is* a scientific account of how a system works (at least a much better one than anything like the magic boxes enterprise). Besides, Searle has obviously never tried to construct a connectionist model without knowing beforehand what the system is supposed to do.

Consciousness Explained?

Moving up from neuroscience to psychology and philosophy, we should not miss the influential (and controversial) work of Dennett (1992). It should be noted that he discussed "consciousness“ as any manifestation of mental activity (that is, the broader philosophical view identified in the introduction). He admitted that his explanation of consciousness is not a "literal, scientific" theory, but rather a collection of metaphors that help understanding better the essence of conscious thinking. Dennett based his theory on ideas from cognitive psychology, AI and philosophy.

Dennett used the view of Baars (1988) that "consciousness is accomplished by a 'distributed society of specialists that is equipped with a working memory, called a *global* workspace, whose contents can be broadcast to the system as a whole' ". This view is similar to the **Society of Mind** approach of Minsky (1985), and looks actually like a paraphrase of the Blackboard System architecture from AI (which hardly contributes much to the credibility of the view.) The significant idea here is that no single process is responsible for consciousness, but rather many streams of consciousness exist.

An important idea that Dennett promoted (attributed to Kosslyn) was that consciousness is a newly-emerged process in the course of evolution and as such differs substantially from the underlying old and fast processes that serve basic functions like grasping, dunking, locomotion control, etc. He argued that conscious activity is slow and serial because the "brainware" it uses is not designed for its purposes; the brain has evolved to support reactive behavior rather than deliberate conscious planning and reasoning. This explanation for the serial character of consciousness is completely orthogonal to the explanation based on a binding mechanism, but nevertheless is no less appealing.

Another merit of Dennett's work on consciousness is that he reviewed most of the prominent approaches to the study of consciousness in a very critical way. For example, he declared Koch and Crick's theory as unfinished (along the line that it is not known to whom the abstracting mechanism presents the "results of the underlying computations").

Similarly, Edelman's theory was qualified by Dennett as an "instructional failure", while the connectionist approach received somewhat better appraisal.

Flanagan (1992) supported the approach of Dennett and proposed a philosophy called by him constructive naturalism. He argued that consciousness is possible not only in humans, but also in animals, under the condition that their brains possess the necessary complexity.

Consciousness Does Not Exist

This is a possible position too, if we treat consciousness in its broad meaning as adopted in philosophy. The main argument in support of this view is that consciousness is just an abstract term that has nothing to do with the processes taking place in the mind of a thinking subject.

Why this is a possible position can be best illustrated by means of an example from the history of science. The nature of life was till the nineteenth century as great a puzzle to science as consciousness now is. The predominant explanation of life was that a special vital force, **vis vitalis**, existed that could somehow bind together inorganic components to form organic matter. Vis vitalis was a very convenient form to explain the difference between living and nonliving matter without going into details.

What replaced vis vitalis today, after organic compound have been synthesized? Nothing but complexity. Organic substances are substantially more complex than inorganic ones, and it is the complexity of organic substances that distinguishes them from inorganic compounds. However, complexity is not endemic only in biological systems - it is throughout the universe. Could this be also the answer to the mind/body problem? Quite possible, though it will probably take a different order of complexity to explain materialistically a phenomenon as difficult as mind and consciousness.

Consciousness might be viewed also as a recent embodiment of the ancient concept of soul (Revonsuo, Kamppinen and Sajama 1994). To Aristotle soul meant something that makes a thing a living thing; that is, a principle of life. What is today considered biochemistry was to Aristotle undoubtedly an element of the soul. Subsequent advances of science elucidated the nature of life and effectively limited the content of the term "soul" to what is nowadays equivalent to "consciousness" in lay language. It can be expected that science will put further pressure on this term until maybe it loses all content.

The people that put an end to the argument about the nature of life were chemists, not philosophers and metaphysists. It can be believed that the people that will solve the mind/body problem will analogously be not philosophers, but neuroscientists and/or roboticists. What it takes is to construct a thinking creature or analyze the operation of one - certainly not as easy a task as the synthesis of organic compounds, but definitely that concrete and probably fully realizable.

The position defended here should not be confused with functionalism and reductionism, which generally reject the existence of mental events (Churchland 1984). Mental events exist and probably should be studied by methods different from those that are used to study the physical brain. The existence of mental states, however, is not a good enough reason to divide objects into conscious and unconscious ones.

Consciousness cannot be explained

This is different from "Consciousness does not exist". What people defending this position mean is that the mechanisms giving rise to consciousness are too complex to be studied with the present analytical apparatus of science. Roger Penrose (1989) contended that the processes responsible for consciousness are nondeterministic and chaotic and as such cannot be studied efficiently by physics (and hence by neural science). This position merely states that there are "technical" difficulties in studying consciousness that are not likely to be overcome.

Penrose and Hameroff claimed that cosnciousness emerges from the operation of the microtubules, tiny tracts built of protein that surround the neural cells. Even though their hypothesis seems fictional, it demontrates why the problem of studying consciousness might be intractable if consciousness is indeed based on some micro-level effect.

Another "technical" limitation that might arise in the quest to understanding consciousness is the cogntive capacity of our own minds. McGinn (1991) pointed out that much like lower animals cannot understand the scientific disciplines conceived by man, man himself might not be able to understand the operation of thinking matter. This is always a viable possibility - certainly thousands of years of scientific inquiry haven't done much to disprove this position.

Conclusion

A controversial topic like consciousness inevitably gives birth to many theories and views, part of which tend to obscure rather than clarify the subject. The position taken in this paper was that most of the approaches from philosophy unfortunately fall into this category. It was argued that consciousness could be studied if its meaning is qualified carefully from the perspective of psychology and methods from neuroscience are employed to gain insight into the neuronal mechanisms that give rise to conscious activity. To support this view, several concrete research paradigms were discussed that show promise in this respect.