

Commentary **Antti Revonsuo**

Research Fellow of the Academy of Finland. Department of Philosophy, Center for Cognitive Neuroscience, University of Turku

Notes on the Foundations of Consciousness Research

Introduction

I fully agree with the basic starting point of the target paper: Psychology should take consciousness seriously, for it is the branch of science into whose domain subjective mental phenomena fall. Behaviorism managed to purge all subjective phenomena from the domain of psychology. Cognitive science did not do much to welcome them back, either. Now the time is ripe for the return of consciousness to psychology, and for a new beginning of a multidisciplinary science of consciousness.

Overgaard says that his goal is not to defend any particular framework in the study of consciousness, but only to discuss the fundamentals that any such framework must take into account. His target paper examines some of the most central topics in current consciousness research: How should the concept of consciousness be defined? How could we find the neural correlates of consciousness, and would they constitute an explanation of consciousness? What is the role of introspection in experiments on consciousness? I also believe that these are among the very central questions in the field of consciousness research. I do not believe, however, that a totally neutral point of departure is possible. As we have learned from philosophy of science, any point of departure is bound to include some philosophical commitments and background assumptions. They may be implicit and one might not be aware of them – it is often said that empirical scientists are not aware of the philosophical background assumptions they make – but even so, such assumptions are inescapable.

It seems to me that Overgaard is to some extent under the illusion that his approach is a rather neutral starting point for the science of consciousness. I will try to point out in my commentary that that is not the case: Overgaard is as deeply committed to certain philosophical ideas about consciousness as anybody else in this field. Although I largely agree with Overgaard about the *problems* that the science of consciousness faces, I am afraid that I do not share the philosophical commitments concerning consciousness with him, and therefore I also disagree about the *solutions* to the problems. Be that as it may, I believe it is useful to point out our disagreements. They are a demonstration that these foundational issues must be thoroughly clarified and discussed, lest we should proceed under the false impression that they have been already solved for good, and that

somehow everyone agrees what we are studying when we study consciousness.

The concept of consciousness

The most fundamental concept in this domain, obviously, is the concept of consciousness. Hence the definition of consciousness is not an insignificant issue, but perhaps the most important task at this stage of development of the field. But we should remember that no definition of consciousness is going to be free of philosophical background assumptions. This is reflected in the current controversies: some define consciousness as *phenomenal* consciousness (subjective experience and qualia); others as *access* consciousness (input-output function of conscious information processing); some deny the existence of phenomenal consciousness and define consciousness in terms of skills and action; some define consciousness as a higher-order mental state, and so on. One's definition of consciousness thus is bound to include a concise expression of one's philosophical theory of consciousness. Different philosophical views lead to dissimilar explanations and empirical research programs on consciousness. Hence, the definition of consciousness has far-reaching implications for the study of consciousness, for it delineates the nature of the *explanandum* of the field: what is to be explained by a science of consciousness?

Let us take a look at Overgaard's proposed definition which is actually a definition of a conscious *state* rather than consciousness *per se*:

A conscious state is a mental state, individuated by its content, a subject is directly aware of being in. The conscious state is had by a subject who is in the state.

The central elements of this definition are "subject", "direct awareness of", and "content". The philosophical background assumptions included in this kind of definition are the following: there is an entity called a "subject" that is independent of consciousness. This entity can "have" conscious states by engaging in a specific relationship (or mental act) with (otherwise unconscious) mental states, the relation of "direct awareness". The direct awareness is "about" or directed at the contents or objects of the state (intentionality). Thus, Overgaard's definition includes the following tacit assumption about the structure of

consciousness: Consciousness is a matter of a subject relating itself to an object (or content) through an act of direct awareness. According to such a definition, the structure of consciousness is inherently bi-polar: any conscious experiencing has (at least) two components: the direct awareness, and the phenomenal contents or objects that are presented to this awareness. In philosophy, this doctrine is known as *the act/object structure of experience* (Stubenberg 1998).

I will not go here in any detail into the many fundamental problems that act/object-theories of the structure of consciousness involve. But just to mention a few of them: What is “direct awareness”? Is it some kind of “pure sensing” that has no qualitative feels in itself? If so, what distinguishes it from nothing at all? What are the “contents” or “objects” of this sensing; are they external physical objects, internal neural states, or internal unconscious mental states? How do they manage to bear qualitative properties, and what exactly happens when the act of awareness touches them and subjective conscious experience proper comes into being; when the subject becomes aware of the qualitative properties?

Although the act/object structure of experience is a widely held (and usually unquestioned and tacit) assumption about the structure of consciousness, it is deeply problematic. Excellent analyses of the problems involved, as well as alternative ways of defining consciousness that reject the “subject-act-object”-structure, can be found in Stubenberg (1998) and Dainton (2000). I personally find the critique of act-object theories devastating, and the alternatives more inviting. Thus, in the following I briefly describe how I approach the task of defining consciousness. In my view the basic structure of consciousness is simple and non-relational. There is no underlying subject having awareness of objects: there are simply patterns of subjective, qualitative experience without any distinction between awareness and contents. The subject is simply the sum of simultaneously present and globally unified experiences rather than some entity external to the experiences that “has” them.

In my view the most basic concept in consciousness research is “primary consciousness” (Farthing 1992) or “phenomenal consciousness” (Block 1995). Its core element is *phenomenality* or *subjective experience*. A sufficient condition for primary consciousness is the mere presence of any type of subjective experience for an organism. Primary consciousness could thus be defined as the current *presence* of subjective experiences. Phenomena in primary consciousness only exist in the form of present subjective experience. The presence of subjective experiences defines what it is like to be this subject, here and now; what it is like for the subject. To have primary consciousness only requires that there are some patterns of subjective experience present (for the organism). It is purely about the existence of any sorts of patterns of subjective experience; whether simple or complex, meaningful or meaningless.

Primary consciousness is a necessary condition of reflective consciousness, another basic concept. When we have reflective consciousness, we deliberately focus on some feature or element of our experience in primary consciousness, i.e. some content in primary consciousness becomes the *object*

of our attention in an act of reflection. The act of reflection can include the questioning, evaluation, comparison or description of the elements found in primary consciousness. To put this in information processing terms: reflective consciousness takes phenomenal consciousness as its *input*, and performs some further cognitive processing (under voluntary control) on that phenomenal experience. Reflective consciousness is closely related to the concept of *introspection*. When we introspect, we use reflective consciousness with the intention to deliberately observe and report the contents of primary consciousness. Primary consciousness includes the mere presence of phenomenal experience; reflection on this experience involves focused attention and an attempt to categorize, conceptualize, evaluate or think about the experience. Introspection on this experience furthermore involves the deliberate attempt to translate some aspects of the experience into a reportable form. The form may be propositional if a verbal report is needed, pictorial if a graphical representation is required, or just a keypress if an explicit categorization of the experience should be reported.

Note that when defined in this way, reflective consciousness and introspection usually involve attending to *external* sources of stimulation rather than to some internal “feels” of “what it is like for me to have this experience”. Attend to the sky (which seems to be out there) and reflect upon its blueness, and you engage in reflective consciousness. Tell about the blueness you see to your companions and you engage in introspection. This is in fact the way in which classical introspectionists also used the concept. Classical Introspectionists (e.g. Titchener 1896) said that when using introspection, we should attend to the stimulus, and by doing so, the sensation becomes clearer:

Be as attentive to the object or process which gives rise to the sensation, and when the object is removed or the process completed, recall the sensation by an act of memory as vividly and completely as you can. The object or process which gives rise to a sensation is termed the *stimulus* to that sensation. If we attend to the stimulus, the sensation becomes clearer, and has a more enduring place in consciousness than it would have gained in its own right. (Titchener 1896, p. 33).

Overgaard however defines introspection in an entirely different way. This is a summary of what he says about introspection:

Reporting about a conscious state involves making the state the object of one’s attention. That is, one could claim that while any other report about objects in the world involves observing and attending to the relevant object “out there”, a report about a conscious state involves observing and attending to the way in which something is experienced. *An introspective state is a mental state by way of which the subject is aware of being directly aware of being in a conscious state.* Being introspective involves an attending to consciousness.

The conflict between the classical way of defining introspection and Overgaard’s definition is serious, for they exclude each other. Overgaard explicitly says that

introspection is NOT the attending to objects experienced “out there”, but attending directly to consciousness itself. Titchener, by contrast, wrote that:

“Direct introspection – observation of a process which is still running its course – is, as a matter of fact, entirely worthless; it defeats its own object. Psychological introspection, however, does not consist in the effort to follow up a process during its course.” (1896, p. 33)

He even gave the following example:

Suppose, e.g. that I am absorbed in the enjoyment of a humorous story or a musical composition, and suddenly ask myself what my enjoyment is, and what mental processes go to make it up. I find myself baffled: the putting of the question has seriously altered my consciousness. I cannot enjoy and examine my enjoyment at one and the same time. (1896, p. 33)

Overgaard says that his starting point for defining introspection is not classical introspectionism, but I doubt that he is aware how radically his definition departs from it. Overgaard’s definition seems to follow from his philosophical background theory (the act-object structure of consciousness). The implicit idea in these theories often is that the “object” or “content” of consciousness is something that is “out there”, whereas the “act of direct awareness” is something “in here”. Overgaard’s introspection aims at capturing the *internal act of awareness*; the experience of consciousness in the process of formation when awareness reaches out for its objects. But that kind of notion of introspection is full of problems both philosophically and empirically. I believe the science of consciousness would be well-advised to follow the classical introspectionist definitions instead and to attend to the *stimulus as experienced* rather than to some sort of elusive internal experience of consciousness as such, independent of the experienced stimulus, which is a metaphysically problematic and methodologically hopeless notion of introspection for empirical science.

“State” of consciousness

Another concept discussed by Overgaard is the “state of consciousness”. Here he gives two possible definitions: Either “state” means exactly the same as “content” (in the sense that any two different contents of consciousness imply two different states of consciousness; only identical contents imply the same state of consciousness). Or “state” means “consciousness as such”, which means consciousness independently of any contents; what is left of consciousness when all contents are taken away; or what is common across all the different contents of consciousness. Again Overgaard’s commitment to the act-object structure of consciousness can be recognized in the background of these notions. Either states of consciousness are individuated together with the objects/contents of consciousness (in which case there are as many different states as there are different contents of consciousness, which makes the concept of state superfluous), or with the act of awareness, which is common to all contents of consciousness and does not vary with content (in which

case there must be some sort of wholly contentless awareness, or “pure” consciousness, without any contents whatsoever).

I find both of these definitions of “state” as problematic as the background theory from which they are derived. I believe that a more viable notion of “state” of consciousness can be defined by starting with the notions, already established in empirical psychology, of “normal state” and “altered states” of consciousness (Kallio and Revonsuo, 2003). The “normal state” is consciousness during alert wakefulness; typical altered states include hallucinations and delusions caused by various changes in the background mechanisms of consciousness (dreaming during REM sleep; hypnagogia; drug states; hypnosis; meditation, etc.). According to this analysis, the “state” of consciousness refers to the *background mechanisms* which *modulate* phenomenal consciousness (the patterns of subjective experience) and *mediate information* from the world and the body to phenomenal consciousness. When these background mechanisms work in such a manner as to produce hallucinatory or delusional contents of consciousness, then the state of consciousness is altered. When they don’t work at all, then no contents of consciousness are possible. Hence, these background mechanisms are shared by all different contents of consciousness, and there are as many different types of states of consciousness as there are different ways in which these background mechanisms can be altered from their normal mode of operation. But these mechanisms are not themselves conscious or parts of phenomenal consciousness, or acts of pure contentless awareness; nor can they be reached or described by introspection. They are the nonconscious neural mechanisms that make phenomenal consciousness possible, and that allow the accurate conscious representation of the world. If we study “states” of consciousness, we study these nonconscious background mechanisms into which phenomenal consciousness is deeply embedded: how they generate or modulate patterns of experience, and what guarantees that the patterns of experience accurately represent the physical stimulus environment instead of constituting a hallucination.

Correlation and Explanation

Another prominent theme in Overgaard’s paper (and in current consciousness research) is the role of correlations between neural and conscious phenomena in the explanation of consciousness. I agree with his conclusion that a correlation theory of consciousness would not provide us with an explanation of consciousness. But I disagree with his claim that cognitive neuroscience simply studies correlations between the brain and the mind and leaves it at that. It is true that the *data* (the actual observations and measurements in cognitive neuroscience) mostly reveal just correlations. The aim of scientific explanation is however to go *beyond* mere data and to present *explanatory theoretical models* which include descriptions of unobservables such as invisible micro-level mechanisms that are invoked to explain the observed correlations in the data. In the biological sciences, explanatory

models (e.g. the cell theory) are not lists of different types of observed phenomena and the statistical correlations between them. By contrast, biological explanations are multi-level mechanistic models that describe how micro-level phenomena and their complex causal interactions can constitute higher-level phenomena. An explanatory model describes a multi-level, causal-mechanical hierarchical network and explicates the causal and part-whole relationships between different types and levels of phenomena. Cognitive neuroscience, along with the other biological sciences, surely aims at these kinds of explanatory models, not mere lists of correlations. Thus, if the question is: "Can cognitive neuroscience explain consciousness?", it is not to be read as "Can brain-mind correlations explain consciousness?", but as "Can a multi-level causal-mechanical model of the brain explain consciousness?". The latter question, I believe, remains to be settled – in fact, discussion around it has hardly even begun (for more on these issues, see Revonsuo 2001a,b,c).

Conclusion

I have only scratched the surface of a few of the many important issues reviewed in Overgaard's paper. A full analysis and exploration of all of them is not feasible within the space of this brief commentary, but must be left for another occasion. In conclusion, I agree with Overgaard about the conceptual and methodological problems that the emerging science of consciousness must face; what I don't agree about are the solutions he suggests to them. I have briefly described my own, alternative solutions to some of the problems. Whether they are any more plausible than those offered by Overgaard I must leave for the reader to decide.

In any case I hope that I have succeeded in pointing out in this commentary that when we define the fundamental concepts of a new field, we must be extremely sensitive to and acutely aware of the philosophical commitments we are making, for they will guide further research long after everybody may have forgotten what precisely the commitments were and why they were made. If we get something wrong from the start, or are unaware about some of our philosophical commitments and their implications, we may be led astray in our empirical research program on consciousness.

If there's any time when the science of psychology should have to be maximally clear about the philosophical commitments it is are making, the time is right now when we are laying the foundations of the new field of consciousness science.

Acknowledgements

Antti Revonsuo is supported by the Academy of Finland (project no 45704)

References

- Block, N. (1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227-247.
- Dainton, B. (2000) *Stream of Consciousness*. London: Routledge.
- Farthing, W.G. (1992) *The psychology of consciousness*. New York: Prentice Hall.
- Kallio S & Revonsuo A (2003) Towards an empirically testable account of hypnosis. *Contemporary Hypnosis*, 20(3): 111-164
- Revonsuo A (2001a) Can functional brain imaging discover consciousness in the brain? *Journal of Consciousness Studies* 8 (3): 3-23.
- Revonsuo A (2001b) Discovering the mechanisms of consciousness. *Journal of Consciousness Studies* 8 (3): 44-50.
- Revonsuo A (2001c) On the Nature of Explanation in Neurosciences. In: Machamer P, Grush R & McLaughlin P (Eds.) *Theory and Method in the Neurosciences*, 45-69. Pittsburgh: University of Pittsburgh Press.
- Stubenberg L (1998) *Consciousness and Qualia*. Amsterdam: John Benjamins.
- Titchener EB (1896) *An Outline of Psychology*. New York: Macmillan.