

Commentary

Andreas Roepstorff

Centre for Functionally Integrative Neuroscience, University of Aarhus

Outlining the Sandpit of Consciousness Studies: a question of Foundations or of Style?

During the 1990's a novel research field called 'consciousness studies' emerged at the interface between psychology, neurobiology, and philosophy. It is not quite clear which driving force gave the development its current thrust. Technology in the form of refined, advanced brain scanning techniques clearly played a role, as did important changes in conceptual models and intellectual interests within psychology and philosophy as well as neurobiology. No matter the causes, a novel terrain has been opened up between the known and that which cannot be known at present. It is this territory that 'consciousness studies' try to colonise, by transforming what is often described as one of the last white spots on the human knowledge map into a thriving disciplinary landscape with established facts, canonical texts and university chairs. At present, however, consciousness studies are very far from that position. It is rather to be likened with a newly opened large sandpit, located in a multi-cultural neighbourhood. It attracts players from all sorts of intellectual cultures, but the rules of the game are not yet fixed, and various principles appear to apply in different corners of the pit. Here and there, journals and organisations with varying degrees of authority, such as *Journal of Consciousness Studies*, *Consciousness and Cognition*, *Association for the Scientific Study of Consciousness* and the annual conferences *Toward a Science of Consciousness* function as integrating nodes that tie together an emerging network. And as with the play that goes on in a sandpit, the intellectual constructions being made seem still transitory and open ended, not yet solidified into that apparently dense structure, supported by text books, educations and curricula, which characterise 'proper' sciences.

Morten Overgaard's work, both in the target article for this issue of *Forum for Antropologisk Psykologi* and his recent Ph.D. thesis, is one of the first systematic attempts by a Danish psychologist to move into this field. As such it puts itself in the interesting position of having to please two different audiences at the same time. One is the fellow psychologists of the particular Danish breed, who by reading the article may be convinced that there is a novel field they ought to pay attention to, despite the wide-spread back-bone reaction against anything that smacks of naturalisation, experimentation and brain. The other audience is fellow participants in the consciousness studies sandpit who, presumably, are eager to see their field upgraded from the rather faint notion of 'studies' to the high-polished structure of 'science'.

I do not belong to the first category of the audience. I am not a psychologist, and although I have done participant-observation style investigations of psychologists and currently work in a brain imaging laboratory with psychologists as some of my closest colleagues, I do not share entirely their intellectual heritage and disciplinary ancestors. It is therefore difficult for me to evaluate how the argument plays itself out vis-a-vis current and historical trends in Danish psychology. My reception is more shaped by the perspective of the second category, it is read by someone who is at the same time embedded in and trying to get an understanding of the consciousness studies sandpit. And from that position, I find it difficult not to be sympathetic to Overgaard's approach. There is a careful discussion of some of the conceptual and methodological problems in consciousness research, from both sides of the 'hard problem' of consciousness (Chalmers, 1996) so to say. That is, the reader is presented with some of the problems both in the third-person 'scientific' – read neuroscience – approach to consciousness and in the phenomenologically inspired insistence on the uniqueness of the first-person perspective of the perceiving subject.

This point becomes highlighted when these different levels of description are brought together in the search for the so-called NCC, the "Neuronal Correlate of Consciousness". The issue at stake is ontological: is there such a 'thing' in the world as a neuronal correlate of consciousness, as well as epistemological: how should one come about getting a 'scientific' – what ever that is taken to mean – grip on the consciousness-brain relationship?

Both of these questions are terribly interesting and, for strangely interrelated reasons, none of them seems answered, perhaps they are even unanswerable at present. However, in spite of these fundamental problems, the discourse in itself seems to be one of the most important arenas for the examination and delineation of human nature at the beginning of the 21st Century (Roepstorff, 1999). In spite of the unresolved epistemological and ontological problems, the discourse is therefore likely to take on a cosmological function (Roepstorff, & Bubandt, 2003), it is also about constructing a grand narrative about humanity and its place within a universe both of meaning and of nature. As also argued by Overgaard albeit in different words ("the current boom in cognitive neuroscience seems so overwhelming that it potentially could have a lasting impact on the way we think of ourselves"), this is one reason why the readers of this journal may take an

interest in the issue, even if they do not care much about the cumbersome details of brain imaging or psychophysical examinations.

It is one of the great merits of Overgaard’s paper that he drags ‘the hard problem’ out of the abstract realm of philosophy and into the practical mess of empirical examinations. This is aptly done in the discussion of the search for NCC. Overgaard convincingly claims that two different types of correlations are present: a ‘desired correlation’ between conscious states and brain states, and an ‘actual correlation’ between a behavioural indication and measures of brain activity (see figure page 26). This aspect is currently underdiscussed in consciousness studies, and many researchers seem to write up their articles without realising that their argumentation – apparently seamlessly – slips back and forth between these two different forms of correlation. However, I think that the analysis proposed by Overgaard may be taken even further. In the following, I will try to do this by 1) renaming the correlations and expanding the figure into two different levels, an ontological and an epistemological and 2) embedding this in a larger cosmological framing.

In Overgaard’s representation of the problem, we get an apparently one dimensional representation of two times two

‘conceptual boxes’ connected with a line; this is overlaid with two correlations each indicated by a one-sided arrow. There is of course no strict rule for reading a diagram like this, but one rather obvious interpretation is a more or less direct link between conscious states and behavioural indications on one side, and between brain states and measures of brain activity on the other. However, this reading is contested by the fact that while the arrow ‘desired correlation’ is dotted, the one for the ‘actual correlation’ is not. How should we, then, read the figure? If the links are unproblematic, the two correlations become homologues; this would render Overgaard’s distinction between the two types of correlation overtly pedantic, even trivial. Overgaard clearly – and I believe rightly – argues against the two correlations as homologues, therefore something ‘fishy’ must go on inside the apparently unproblematic links, but we do not get much help as to what that could be. As for the first link, Overgaard claims that one must “accept a relation between reports about conscious states and the conscious states themselves as an *a priori*” while the second link – between brain states and measures of brain activity – is not much discussed at all. I think this interesting tension calls out for an expansion of Overgaard’s analysis (Figure 1)..

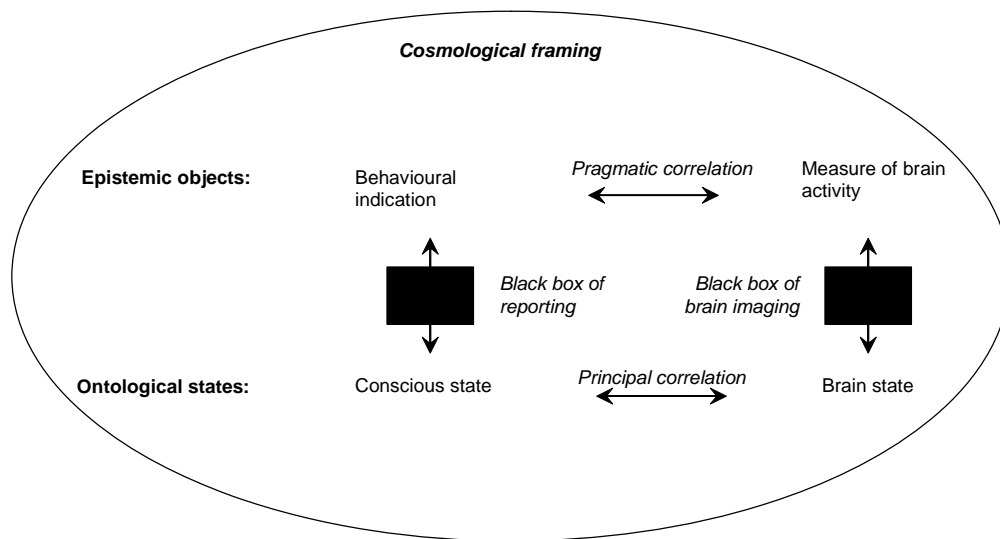


Figure 1: The fact making square of consciousness studies

I propose to locate the two correlations discussed by Overgaard at two different levels. Ontologically speaking, it does make sense to speak of the existence of a particular conscious state and of a particular brain state. However, as rightly argued by Overgaard, in the actual practice of conducting experiments and writing up papers, the correlation takes place between behavioural indications and measures of brain activity. Now, there clearly are such things in the world as measures of brain activity and behavioural indications – either in the form of verbal reports or button presses – but their mode of existence seems somewhat different from that of a conscious state or a brain state. They are rather to be understood as *epistemic objects*, that is, as the outcome of a particular set of methods and criteria that ensure some kind of validity, ideally approved by the intersubjective evaluation of a particular scientific community. Within science studies, it has become customary to describe this process as a ‘black boxing’, this means that when it runs smoothly “one need focus only on its inputs and outputs, and not on its internal complexity” (Latour, 199, 304). Overgaard is effectively ‘black boxing’, when indicating the straight lines in the figure (p 28). However, as has been demonstrated by a whole range of science studies, the relation between the resulting epistemic objects and the underlying states is in no way trivial. The commonly applied counter strategy has here been to ‘open the black box’ in order to follow in minute details the actual transformations, reductions and amplifications involved in settling the epistemic objects (see Roepstorff, 2001b for an extended discussion). I will in the following shortly outline some of the processes that Overgaard is black boxing.

Although it may not be generally known outside the brain imaging community, it is relatively uncontroversial that the colourful pictures of brain activity obtained by PET, fMRI, MEG or EEG are very far from realistic photographs of the brain. They are rather to be seen as complicated graphs, the outcome of a set of mathematical procedures and transformations, that could have been done differently (Roepstorff, 2002, 2003b; Roepstorff & Gjedde, 2003). To complicate things even further, there are serious discussions in the brain imaging field about what the relationship is between the largely metabolic and circulatory measures obtained and the actual behaviour of neurons. These discussions occur at two levels. It is a matter of settling the link between, for instance, fMRI measurements of the BOLD signal or PET-measurements of blood flow or oxygen consumption and the underlying neuronal activity. More fundamentally, however, there is not agreement as to what should count as a proper description of brain states, should they, for instance, be identified by synaptic processing or by the firing of neurons? This means all in all that the link between the measure of brain activity and the putative brain state is in no way trivial; and it is stabilised as much by conventions that are still changing as by uncontested quantifications.

I believe that a parallel argument can be put forward for the link between behavioural indication and conscious states: the resulting epistemic object is also the result of a process of black boxing. This point may be slightly more controversial, and this is not the place for a full argumentation (see however, the discussion in Jack & Roepstorff, 2003), but

two examples may suffice to briefly outline the problematique. One line of evidence is demonstrated in an interesting set of experiments pioneered by Tony Marcel (1991). They suggest that even in a very simple psychophysical setting, the behavioural measurement depends critically on the actual method of reporting, be that button pressing, verbal account or eye blink (see also Marcel, 2003 and Overgaard’s dissertation for further discussion). Another line of evidence is suggested by the claim that ‘reports’ be that button presses, questionnaires or unstructured interviews are the result of intersubjective processes with all that it entails in terms of sharing or non-sharing of trust and frames of reference (Jack & Roepstorff, 2002; Roepstorff, 2001a, 2003a; Roepstorff & Frith in press). Both of these examples can be interpreted to suggest that the relation between the reports and the behavioural measurements on one hand, and the putative underlying conscious states on the other, is not trivial. As with the relation between brain images and brain states, it is the result of the application of particular epistemic technologies, and only through a careful interplay between black-boxing and opening the box does it become possible to elucidate this relation.

This way of setting up the scenario is a further complication because the apparently unproblematic links in Overgaard’s model have been replaced with black boxes. This is not trivial, because it suggests that the situation is much more open than indicated by Overgaard. It is not only a matter of realising that the ‘actual correlation’, which one may experimentally validate, is different from a ‘desired’ relation between ontological states. The problem is also that we really don’t know what the relation is between these ontological states, whose existence is hypothesised rather than proven, and those somewhat problematic epistemic objects, that are the result of certain methods and techniques. Rather than talking about ‘actual’ and ‘desired’ correlations, notions that carry almost Freudian semantic resonances of a paradise always already lost, I would therefore prefer to talk of ‘pragmatic and ‘principal’ correlations. Whereas Overgaard seems to argue that these correlations go from the domain of mind (consciousness, behavioural indications) to the domain of the brain as indicated by the direction of the arrows, I don’t think that is the only possible trajectory. It is rather a matter of establishing a process of *triangulation* that symmetrically attempts to bring together different forms of epistemic objects i.e. subject reports, behavioural measurements and brain images in a joint investigation of underlying states of consciousness as well as of brains (Jack & Roepstorff, 2002). This entails realising that the way the black boxes are configured have consequences both for the construction of the epistemic objects and for the constitution of the ontological states they are allowed to refer to. Furthermore, once a particular set of black boxes are well established, and the fact making square indicated in figure 1 is up and running, the contours of a particular cosmology are already more than in the making. Skinnerian behaviourism with its focus on third person descriptions constructed certain types of black boxes that allowed for one cosmological configuration (Baars, 2003), the current move to include consciousness is likely to make another construction, but that particular design will be

different depending on whether focus is on the uniqueness of the first person experience, the current trend in consciousness studies, or on an interplay between first person experience, third person description and second person interactions (Roepstorff, 2001a).

Now, the realisation that this is, indeed, along several dimensions an open examination of a novel field allows for an understanding of what it means for an examination to be scientific that has a somewhat different focus than proposed by Overgaard. Overgaard seems to focus on 'deep elements', that is, on defining principles and identifying foundations. This is scientific in the sense that it is about order and structure, and this is the determination of science usually favoured by arm-chair philosophers. But scientific examinations are not only characterised by a call to order. It is also a matter of playful intervention, of curiosity, of interest in a particular topic motivated by a desire, also in the Freudian sense, to 'find out' and to explore that constantly shifting terrain of potentials which is located between the known and that which cannot be known at present.

In this understanding of a scientific field, which is favoured by many people who actually study the sciences empirically, like anthropologists, sociologists and historians, it is not so much 'deep principles of foundation' that counts. Rather than 'deep' coherence, a scientific field may be characterised by institutional settings on one hand, and by a somewhat more shallow notion of 'style' (of knowing, thinking or reasoning), at least in the sense where this notion has been applied by historians and philosophers of science like Ludwik Fleck and Ian Hacking. To paraphrase the former, a style of thinking may be characterised by common features in the problems of interest, by the judgement which the thought collective considers obvious, by the methods which it applies as a means of acquiring knowledge, and it is usually accompanied by a technical and literary style characteristic of the given system of knowledge (Fleck, 1979; see Roepstorff, 2003b for a discussion).

From that perspective, consciousness studies is currently characterised by an intense exploration of stylistic possibilities. Let's try to go through them one by one. Is consciousness an interesting problem? Well, if one doesn't think so, then there are probably better sandpits to play around in elsewhere. Does it make sense to use well-known psychological entities as awareness, emotion, memory, anticipation and experience as variables in examinations of the brain? Probably, it seems at least that interesting, somewhat consistent results may be found when it is attempted, although there are all sorts of foundational problems related to how one validates and interprets paradigms of this kind (Roepstorff & Gjeede, 2003). Are whole brain neuroimaging methods such as PET, fMRI, EEG and MEG useful means of acquiring knowledge about the brain? Indeed, although the physiological interpretation, let alone the functional ramification, is still not well-established. But it seems that careful and precise experimentation using psychological entities as variables may give some of the clues. And what about introspective and experiential evidences, can that tell us anything about the mind? Well, the answer is both yes and no. Studying consciousness without taking consciousness into

account seems a little ridiculous; at least it is worrying for the resulting theoretical constructs. However, the question as to how this should count as evidence, what epistemic objects should be allowed, is not settled at all (Jack & Roepstorff, 2003). Can this tell anything new about the mind? Open question, at times the brain seems to possess Solaris-like qualities in that the examinations project back to the researcher phantoms of his or her most private dreams (Lem, 1961), at other times, however, independent stability and resistance to interpretation seem to manifest itself. How should one write up these findings? Here a certain literary pluralism is becoming the norm, the researcher aiming to be successful appears to have to write both within several of the strict genres demanded by the most prestigious scientific magazines, and engage in the somewhat hand-waving, speculative set of connections made possible by the novel-like genre of popular science books and magazines.

I think that it is this emerging process of 'the settling a field' which is at stake in the current version of consciousness studies. The constructive work ongoing at present seems to be more driven by curiosity than by principles, more prone to establish novel facts – understood as temporary markers for a resistance to wild speculations (Roepstorff, 2002) – than to search for foundations; and this is probably a healthy sign. I believe that it is very much in the delineation of a community of researchers – who may not share interests completely, may not have same nagging questions about foundations, may disagree somewhat as to what count as a proper argument but who maintain overlapping circles of interest – that the future version of 'what a study of consciousness should mean' will emerge. And it may be that a certain amount of accommodating openness is needed to ensure that the framing of the figure becomes acceptable: that the resulting cosmology becomes both sufficiently coherent and extensive to accommodate and be acceptable to various intellectual traditions. That was something which behaviourism, one of the last attempts to make the study of man scientific, failed to do: the foundations may have been coherent, but the resulting cosmology simply sucked as it was unable to accommodate important aspects of human life (consciousness, intersubjectivity, Roepstorff, 2003a) as well as the researchers who found these issues important.

I think that in this process of establishing both a field of research and of researchers, a reflexive examination of styles of knowing, their resulting facts and the delineations that they allow and enforce may be at least as important as *a priori* attempts at foundational certainty. Of course that preference may precisely be a matter of style; of whether the interest lies in reflexively playing in the pit while examining the contours both of it and the play itself, or whether one is obsessed by trying to find out what is at the bottom (do we hit solid ground if we dig deep enough, or is everything floating on liquids and melted substances?). For now at least, there seems to be enough to do for both strategies to be viable.

References

- Baars, B. J. (2003). The double life of B.F. Skinner. *Journal of Consciousness Studies*, 10(1), 5-25.
- Chalmers, D. J. (1996). *The Conscious Mind. In search of a fundamental theory*. Oxford: Oxford University Press.
- Fleck, L. (1979). *Genesis and Development of a Scientific fact*. Chicago: Chicago University Press.
- Jack, A. I., & Roepstorff, A. (2002). Introspection and Cognitive Brain Mapping: from Stimulus-Response to Script-Report. *Trends in Cognitive Sciences*, 6(8), 333-339.
- Jack, A. I., & Roepstorff, A. (Eds.). (2003). *Trusting the Subject*: Imprint Academics (also published as special issue of Journal of Consciousness Studies).
- Latour, B. (1999). *Pandora's hope : essays on the reality of science studies*. Cambridge, Mass.: Harvard University Press.
- Lem, S. (1961). *Solaris*. New York: Faber & Faber.
- Marcel, A. J. (1991). Slippage in the unity of consciousness. In G. R. Bock & J. Marsh (Eds.), *Experimental and Theoretical Studies of Consciousness* (pp. 168-180). Chichester: John Wiley.
- Marcel, A. J. (2003). Introspective Report: Trust, Self Knowledge and Science. In A. I. Jack & A. Roepstorff (Eds.), *Trusting the Subject*: (pp 167 – 86). Exeter. Imprint Academics (also published in Journal of Consciousness Studies, 10 (9-10)).
- Roepstorff, A. (1999). Det Neurale Menneske- et antropologisk perspektiv. In O. Høiris & H. J. Madsen & T. Madsen & J. Velle (Eds.), *Menneskelivets mangfoldighed*. Moesgaard: Aarhus Universitetsforlag & Moesgaard Museum.
- Roepstorff, A. (2001a). Brains in Scanners: An Umwelt of cognitive neuroscience. *Semiotica*, 134(1/4), 747-765.
- Roepstorff, A. (2001b). *Facts, Styles and Traditions. Studies in the ethnography of knowledge*. Unpublished ph.d. dissertation, Aarhus University, Aarhus Universitet, Afdeling for Etnografi og Social Antropologi.
- Roepstorff, A. (2002). Transforming subjects into objectivity. An ethnography of knowledge in a brain imaging laboratory. *FOLK, Journal of the Danish Ethnographic Society*, 44, 145-170.
- Roepstorff, A. (2003a). A double dissociation in twentieth century psychology? *Journal of Consciousness Studies*, 10(1), 62-67.
- Roepstorff, A. (2003b). Reflections on a brain imaging community. In R. S. e. a. Frackowiak (Ed.), *Human Brain Function*, 2. ed. (pp. in press): Elsevier.
- Roepstorff, A., & Bubandt, N. (2003). Introduction: The critique of culture and the plurality of nature. In A. Roepstorff & N. Bubandt & K. Kull (Eds.), *Imagining Nature. Practices of cosmology and identity* (pp. In press). Aarhus: Aarhus University Press.
- Roepstorff, A., & Frith, C. D. (in press). What's at the top in the top-down control of action? *Psychological Research*.
- Roepstorff, A., & Gjedde, A. (2003). Subjektivitet som variabel i hjerneskanningsforsøg. In D. Zahavi & G. Christensen (Eds.), *Subjektivitet og videnskab. Bevidsthedsforskning i det 21. århundrede* (pp. in press). Roskilde: Roskilde Universitetsforlag.