

On the theoretical and methodological foundations for a science of consciousness

Journal of Anthropological Psychology

No. 13, 2003

Department of Psychology

University of Aarhus

Abstract

The target article suggests a way to define consciousness, and discuss why consciousness has been conceived of as a problem to scientific investigation. Arguments that consciousness cannot be studied empirically are analyzed and refused. From here, the article goes on to suggest a way to perform empirical studies. Introspection is suggested as a sine qua non for consciousness studies, and he discusses with which validity introspective techniques can be applied in experimental settings. The target article reviews the notion of 'neural correlates of consciousness' to explore the explanatory value of such brain measures, and with which strategy such measures can be obtained.

Target-article

Morten Overgaard	<i>On the theoretical and methodological foundations for a science of consciousness</i>	2
------------------	-----------------------------------------------------------------------------------------	----------

Commentaries

Bernard J. Baars	<i>The mind-body problem is scientifically misleading and perhaps useless</i>	14
Antti Revonsuo	<i>Notes on the foundations of consciousness Research</i>	16
Andreas Roepstorff	<i>Outlining the sandpit of consciousness studies: a question of foundations or of style?</i>	20
Erik Schultz	<i>Brain, mind and consciousness</i>	25
Logan Trujillo	<i>Are verbal reports all we will ever have in a science of consciousness?</i>	28
Søren Willert	<i>Is consciousness a thing or an adjunct?</i>	31

Reply

Morten Overgaard		34
------------------	--	-----------

Morten Overgaard

Center for Functionally Integrative Neuroscience, Aarhus University Hospital & University of Aarhus

On the theoretical and methodological foundations for a science of consciousness

Introduction

The scientific exploration of consciousness is presently a popular enterprise within several academic disciplines. Especially, philosophy of mind and cognitive neuroscience have joined forces in the attempt to find a common ground to understand consciousness conceptually and empirically. Now, one might ask, why should psychology care about this development? And why should not just the sub-disciplines such as neuropsychology and cognitive science care, but why should a psychology oriented towards sociology, anthropology or the humanities in general care? One answer is that consciousness is such an obvious feature of our minds: consciousness exists as an intrinsic characteristic of our everyday experience. Behaviourism is a good example of the kinds of problems one gets when leaving consciousness out of one's theory: apart from being trivially wrong when ignoring or even denying consciousness as an empirically observable phenomenon, behaviourism and similar approaches leave the subject with an account of him- or herself that is incompatible with and useless for everyday self understanding.

Now, one can easily argue that psychology needs to be interested in consciousness, and very few would be willing to go against such a claim. One can also, quite easily, argue for the general relevance of the approach of the analytical philosophy of mind in its own terms, trying to define consciousness and related concepts precisely. Also, the empirical attempt to find which cognitive or neural processes that correlate with consciousness seem, *prima facie* at least, relevant within those disciplines. Both approaches can be said to be trivially relevant, for why should one go against a defining of consciousness or reject information about psychophysical correlations. One could however imagine a general psychologist, especially one with a theoretical background in the human or social sciences, think that *even though* a study of consciousness is relevant to psychology, this specific philosophy-cum-neuroscience-approach to consciousness is completely irrelevant to general psychology. In fact, it may be considered reductionistic or very narrow-minded in its understanding of which phenomena to investigate in the exploration of human consciousness, especially.

However, against this view, one can argue that the aim of a "science of consciousness", potentially, is exactly

the opposite: to introduce concepts such as "consciousness", "self", "subjectivity", and "intentionality" into a framework of natural science. That is, given that one can establish a coherent "science of consciousness", one would not simply discuss behaviour, black boxes or brain processes in experimental psychology, but make use of a conceptual framework that opens up the laboratories to general psychology. Ideally, this would clear the way for a cognitive neuropsychology that is not closed around its own logic and ways of asking questions, but one that is able to address issues recognisable in people's everyday life.

The current boom in cognitive neuroscience seems so overwhelming that it potentially could have a lasting impact on the way in which we think of ourselves, just as previous scientific developments have like those of Darwinism or Freudianism. Whereas the Darwinist movement made us conceive of ourselves as animals that came into existence through evolution and the Freudian movement made us think of we as greatly influenced by unconscious desires, a cognitive neuroscience could have just as important and lasting effects. So, it seems likely that we will have different accounts of ourselves when working with a cognitive neuroscience that describes us as conscious organisms and when working with one describing us as, basically, machines or computers. A "psychological" neuroscience instead of a classical cognitive neuroscience, so to say.

So, it shall here be argued that the debate over consciousness that took its beginning in philosophy of mind and neuroscience can be seen as a project with much relevance to general psychology given that 1) what we understand by and can recognise as consciousness can be conceptually grasped, and 2) this understanding can be operationalised and thus be investigated in the empirical sciences (e.g. psychology). Finally, it would be desirable for general psychology if this could lead to 3) a non-reductive cognitive neuroscience that was not in opposition to or irrelevant for our every-day understanding of ourselves.

The concept of consciousness

While scientists have gathered data they claim shed light on consciousness, there has not yet been developed one consistent framework for the gathering and interpretation of such data. If the explosion of interest in consciousness is

going to have long-lasting consequences for psychology and related disciplines, it seems necessary that such a framework is established. Furthermore, if the ambition of several researchers – to form a specific science of consciousness – is to be realised, this seems even more crucial.

The analysis ahead is about principal and methodological issues in the establishing of a scientific approach to consciousness. The purpose is not to defend any specific scientific or theoretical framework about consciousness, such as the “biological naturalism-approach” (Revonsuo, 2001), the “global workspace theory” (Baars, 1988) or the “introspective physicalism approach” (Shallice & Jack, 2001). Rather, the purpose is to discuss some of the fundamentals in any such framework.

Let me begin with the definition of the concept of consciousness that is used in the present context. When attempting a definition of consciousness it is sometimes said that consciousness is identical to experience, or that consciousness is a set of phenomenal feelings (Block, 1995; Searle, 1992). However, all such attempts must face the task of providing a further definition of experience or of phenomenal feelings. Obviously, if such things are to be defined as “consciousness”, we have a circular statement, if they are defined differently, the definition of consciousness is incomplete. Therefore, I shall stipulate the following definition of a conscious state:

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

This definition appreciates the many contemporary attempts to capture the concept of consciousness by reference to the experiential quality of sensations, and sometimes of thoughts and feelings (“qualia”), subjectivity, or intentionality (the “aboutness” or inclusion of the experienced object as the contents of consciousness). It also implies a necessary relation between them: the conscious state is *had* by a subject who is *in* the state. The subject is directly aware of being in this state. Cases in which a subject is told by others that he has some kind of mental state do not qualify in capturing the concept of consciousness – the subject himself must be *directly* aware of it. This notion of direct awareness is not new, but can be found in classical writings of John Locke and David Hume, although I do not consider the definition in any way tied to their philosophical positions. Finally, this state is individuated from other states by its content, which it is about.

To further clarify the definition, one can distinguish between two interpretations of it. A weak reading of the definition could infer that one is conscious if one is directly aware of the contents of one’s mental state. Here, this is what is meant by being *in* the state. The notion of *direct awareness* implies that the subject is better acquainted with the relevant mental state in a fundamentally different way than he would be with the mental states of others. This interpretation expresses the point discussed below that the “seemings-is distinction” is dissolved in the case of consciousness (page 5).

A strong reading of the definition could infer that something further is demanded, namely *knowledge* about the fact that the subject is in a specific mental state. That is, the presence of a thought about the state as such.

One’s choice of reading has important bearings on one’s classification of a number of psychological phenomena. One could take the case of dreaming. Dreaming is, according to the first reading, conscious. According to the latter, it is not. Only in the case of the somewhat rare phenomenon of “lucid dreaming”, people are aware of the state they are in (or at least: aware of dreaming).

The definition is to be understood in the first sense. For one thing, I distinguish between consciousness and introspection (page 12). Furthermore, peoples’ reports (and my personal experience) tell me that there is a *felt* difference between dreaming and being in dreamless sleep. If this difference were not reflected as a difference between consciousness and unconsciousness, we would need new concepts to account for it.

If one accepts this conceptualisation of consciousness as a usable point of departure for an understanding of the concept of consciousness, it should be underlined that it is work in progress. For instance, it is questionable whether “state” is the correct expression for our experiences; some tend to prefer to think of our experiences as a *stream* of consciousness (Varela, 1999), as originally formulated by William James (1890).

The definition, along with most other suggested definitions in the literature, will however give rise to more immediate problems regarding a scientific exploration of consciousness. How, for instance, should a phenomenon, that is essentially subjective and which one subject only has access to, be studied objectively from the perspective of the external scientist?

In a number of different variations, such arguments have been put forward to state that consciousness cannot be approached empirically for reasons of principle: consciousness as phenomenon is beyond the reach of science. Following is a discussion of such arguments.

Conscious states are subjective – science is objective

It is a quite common conclusion for philosophers going against metaphysical positions of a so-called monistic persuasion¹ that the very idea of a scientifically based approach to conscious experience should be met with scepticism. This can be traced historically to a specific understanding of scientific facts – a so-called physicalist understanding, arguing that all things in existence are physical. According to Galilei’s first criterion of what can be

¹ A philosophy saying that only one ontology is needed to explain all phenomena in the world. Consciousness and all (other) physical phenomena should according to this idea be compatible with each other in a common scientific framework.

called science, a scientific object is an object that can be studied from a third person perspective. From this can be deduced either that 1) consciousness is a “tangible” physical object, or 2) consciousness cannot undergo a scientific analysis (it is simply not the kind of thing that can be studied in a third person kind of way). If one accepts physicalism, one can also accept suggestion 1. However, there is at present no consistent suggestion of how to reduce qualia, intentionality and subjectivity². If one does not accept monism, one is forced to accept suggestion 2.

The problem can be stated as a problem of subjectivity versus objectivity: subjectivity is defined as the condition of being bound to a single perspective. Objectivity is captured in the title of Thomas Nagel’s book *The View from Nowhere* (1986), according to which the ideal perspective for natural science is one detached from any perspective – that is centreless and subjectless. So, the problem appears, how can there be an objective understanding to something subjective in nature? The argument can be stated, according to Ronald Chrisley (2001), that the essential subjective nature of experience prevents its full inclusion in an objective worldview.

Nagel’s own answer to this is, however, that one can have a centreless objective phenomenology in which experiences are included although each subject’s perspective is only “one manifestation” of the mental (Nagel, 1974). Nagel himself admits the paradox of trying to have a centreless description of subjectivity, and, in 1986, he concludes that mental and physical objectivity cannot be integrated in a common framework. While such a view makes a science of consciousness possible (hence the objective side of phenomenology), it leaves little hope for an integrated framework with neuroscience and the cognitive sciences.

One can, however, obtain a less pessimistic view by going against the notion of scientific observation as a view from nowhere, and adopting Chrisley’s (2001) suggestion that scientific observation is a view “from anywhere” (inspired by Brian Cantwell Smith, 1996). Smith argues that “the ability to make the world present, and to be present in the world” is dependent on a subject’s being in a particular place. So, one might say that subjectivity is a necessary precondition for “objective” observations. Why is this more than just another paradox? Because if all observation – scientific as well as unscientific – is dependent on a perspective and a subject, then all reports refer to an observer’s experiences. Or, put another way, if one accepts a scientist’s report about, say, observed isotopes in a rock as being a scientific report, one seems forced to also accept reports about conscious experiences as scientific. The only argument against this, it would seem, would be to claim that reports about experiences are somehow different from reports about perceived objects³. But even if this were the case, it would make a science of consciousness more easily

compatible with cognitive science and the neurosciences as these sciences at some level make use of subjective reports, although with a somewhat different methodology. Furthermore, with this approach, one would say that the gathering of observations from subjects about their “conscious events” and the collecting of data from, say, the brain scanner are based upon the same basic criteria for observation.

In a thought experiment, Frank Jackson describes the case of Mary, who is raised in a black and white environment where she learns “everything there is to know” about the neural substrate for colour vision (Jackson, 1986). However, when she ventures outside for the first time, she learns something new and different: what it is like to see the colour red (which she of course was never exposed to before). Thus, Jackson claims there is more to seeing red than physical and functional facts, and physicalism is false when trying to account for the mind. It is assumed by Jackson that Mary has a complete scientific understanding of the brain while not knowing anything about the experience of red.

Chrisley (2002) has raised the argument as a critique of Jackson that the experience of red is part of a complete scientific understanding rather than opposed to it. One may however put forward the counter-argument that while the experience of red is a necessary part of a complete scientific understanding of vision, it might not be a necessary part of the understanding of the brain. This is dependent on one’s metaphysical understanding of the relationship between mind and brain. Obviously, since Jackson goes against physicalism, he would not put forward the claim that a scientific understanding of the experience of red is dependent upon the possibility of deducing the experience of red from knowledge about the brain.

But let me as clear as possible about this issue: what is incommensurable here is *not* science and consciousness, but physicalism and science. Physicalism excludes consciousness when not being able to reduce it to physical facts. Science, on the other hand, does not exclude consciousness. According to my definition, consciousness is directly observable. Given that one does not insist on a “view from nowhere”-notion of scientific observation, consciousness is a “scientific fact”.

I hope hereby to have shown that the idea that one is left with the two alternatives to either deny the reality of irreducible aspects of consciousness or to conclude that they are out of range of science is true only insofar that science is tied to a physicalistic philosophy. Given that one does not endorse physicalism, one has no reason to endorse the subjective-objective dimensions as critique against a science of consciousness either.

Consciousness is as it seems

Chrisley (2002) has argued that in all sciences there is a clear distinction between the studied object “as it is” and as it appears to the researcher. Accordingly, it is a part of the materialist ontology that when one looks away, the object is

² In this context, this is a mere assertion. In Overgaard (2003), this is argued in more detail.

³ This specific possibility is discussed in Marcel (1993) and Overgaard, Kauffmann & Ramsøy (2001)

still there “as it is” while the appearance of it ceases to be. This is however not the case in consciousness studies where the object is the appearance (and vice versa). That is to say, if I were to “look away” (say, go into dreamless sleep, fall into coma etc.) the object will no longer exist. Chrisley has argued that this difference in itself is a serious threat for consciousness to be perceived of as an object for empirical science.

I believe however that this conception is partially misunderstood, and, in the one case where it is not, it is no real problem. First, a distinction between consciousness “as it seems” and consciousness “as it is” is, in a way, upheld in a science of consciousness. Even though the conscious experiences of other people cannot be observed from the outside, they can make reports about them, and those reports form the data of a science of consciousness. So, the reports are the “seemings” and the subjective experiences are the phenomena as “they are”. Only in one case is it not so – the case of one’s own conscious state. So, one can ask, is it so that Chrisley’s argument is sound, while only being applicable to cases in which scientists wish to study their own conscious states?

Again, I do not think that this is the case, though others might disagree. A distinction between the ways an object appears and the way it really is, is not a desirable situation, but a situation that scientists have learned to deal with through a refined methodology. Accordingly, a case in which the scientific object is identical to the way it appears to the scientist is a highly desirable situation. This would make one’s own consciousness not only a possible scientific object, but a “better one” compared to other scientific objects.

Science cannot solve the mind-body problem

Even if one can speak of mental objectivity and in this way accept that a scientific approach to consciousness is not overthrown by Galilei’s first criterion, it could be argued that the project as such is futile. By adding scientific methodology to the philosophical project of finding a logically consistent metaphysical position to account for the mind-body problem, nothing is gained. When one finds statistical correlations between the activation of certain brain areas and certain states of mind, it could still be argued that those areas are identical to the states of mind, or that it is through those areas that mind is realised, interacting, etc. In this way, scientific knowledge does not seem to do us any good.

Wilfred Sellars (1969) argues that any explanation has a hierarchical structure: at the top, we have an unexplained explainer, at the bottom we have an explained unexplainer, and in between, we have explained explainers. That is to say, any argument has a point of departure where certain things are taken as *a priori*. This point of departure is unexplained – no argument is given why it is so. However, it is used to explain other things – that is, certain other things can be logically deduced from the unexplained explainer.

Those are explained explainers that again allow for something else to be logically deduced from them. In the end, we find, as a conclusion, things that can be deduced from the explained explainers but that are not used to explain something else.

In effect, it could be argued that the only *scientifically* grounded theory about the mind-body problem is what Irving Krakow has called the “Correlation Theory of Consciousness” (2002). According to the correlation theory, we will have a list of all possible conscious experiences (or what is sometimes called “private indexicals” in the philosophical literature), a list of all possible brain states, and a list of what can be measured as statistical correlations between them: what is present on list 2 when something specific is present on list 1? This is all there is, and all there can be, to a scientifically grounded theory of consciousness in that any metaphysical claim about this correlation – that is, any explanation of why the correlation is there – demands a further ontological claim. Such a claim is not – at least not presently –verifiable (or falsifiable) by science.

This theory is in no sense a solution to the hard problem or a “bridging” of the explanatory gap. It does not allow one to deduce list 1 from list 2 or vice versa. This, however, makes a science of consciousness very much on a par with cognitive neuroscience that is studying brain-mind correlations in more and more details, although cognitive neuroscience for historical or other reasons rarely speaks of consciousness.

Why would it be impossible for a correlation theory of consciousness to solve the hard problem and explain consciousness based on brain activity? Simply because using private indexicals and asking subjects what they experience while monitoring their brain activity is an unavoidable methodological requirement to collect the correlations. So, deducing list 1 from list 2 that is already mapped onto list 1 is obviously a circular enterprise. Furthermore, excluding the correlations on list 1 from a scientific vocabulary would be just as logically impossible, in that the elements on list 2 are identified and considered interesting only in virtue of the fact that they correlate with list 1. The visual system, is only identified as a system, with specific brain cells as its components, because specific cells are activated when subjects have visual experiences.

Unlike Krakow, I would not hereby conclude that the correlation theory of consciousness is all there is to be said about the mind-brain problem. One could easily say that, while mind-brain correlation is all that presently can be done from a scientific point, philosophy could still debate metaphysics. In fact, it is most likely that an interest in mind-brain correlations is derived from an interest in the (at least) potential theory of the mind’s relation to the body that one may create. Furthermore, to agree with Krakow here, would also be to conclude that no part of the metaphysical positions in philosophy of mind could be operationalised or reformulated as an empirical question. Though this may prove to be the case, it would be highly premature to take it as proven.

Consciousness cannot be deduced from scientific facts

The philosopher Colin McGinn has developed a line of argumentation that could be taken as going against a correlation theory of consciousness. Basically, it goes as follows (McGinn, 1989, 1991):

Premise 1: Phenomenological, introspective or any other possible “first person method” can never reveal what brain properties underlie conscious experience.

Premise 2: Neuroscience can never reveal this either. Although we potentially could find neural correlates of consciousness, we will never be able to understand *how* specific “brain events” underlie experience.

Premise 3: Of course, explanations are not only to be sought in what can be observed, but also in what could be theoretically inferred to give meaning to observations. Inferring concepts in explanations is subjected to a constraint of homogeneity: psychological concepts should not be introduced to explain physical ones and vice versa. To explain “brain facts”, we only need to refer to other “brain facts”, and since consciousness is not observed in the brain, there is no need to infer theoretical accounts of the brain to explain consciousness. There is simply nothing there to be explained. The same conclusion derives from the first person perspective: we do not observe “brain phenomena” introspectively, and thus, nothing forces us to refer to those in an explanation. Due to the homogeneity constraint, this simple fact should prevent us from the very attempt of making explanations of this kind.

Conclusion: You cannot derive mind from brain, brain from mind, or a common structure behind the two from any of them, and thus the mind-brain problem is unsolvable.

This argument contains some of the most often used objections against a science of consciousness. On the surface, the argument seems to have no important consequences for a science of consciousness if one adopts an attitude of a correlation theory. That is, even if one would go along with McGinn’s argument, it would have no direct methodological effects on a framework in which one is only correlating reports about conscious states and measures of physical events. On the contrary, premise 1 and 2 of the argument would only challenge a framework in which a specific kind of causal relation between mind and body is assumed.

On the other hand, premise 3 of the argument does raise doubts about the general point of a correlation theory. If we only need to refer to psychological concepts to explain other psychological concepts, the fundamental idea of even a non-reductionistic interpretation of cognitive or neuroscientific approach to consciousness is simply meaningless.

If one takes a closer look at the argument, McGinn’s first and second premise seem quite uncontroversial. They simply point at the classical epistemological interpretation of the mind-body problem. Premise 3, which logically leads to the conclusion that the mind-body problem is unsolvable in both scientific and theoretical frameworks, is however highly

debatable. If two phenomena, like mind and brain, should be compatible yet do not seem to be so, our whole conception of “mind” and “brain” may be false. So pointing out that two *theoretical* constructs are incompatible does not lead to definitorial unsolvability, even if we were to agree with that claim. It could just as well lead to a suspicion that something in our definition has gone awry.

McGinn’s demands to an explanation seem to be too strong. Even in the cases where science has developed its most exact descriptions and where no ontological problems are conceived, it would be hard living up to this standard. Using the often-used example of water, there is no way of going from our every day perception of water as being fluent, transparent etc. and to the molecular structure of H₂O, and vice versa. Yet, one might argue, we do not conceive of this as a scientific problem. Instead, we believe that there is an absolute “relationship” between H₂O and the qualities of water (in fact so absolute that the term “relationship” seems unnecessarily cautious), where the two epistemological levels give us different kinds of information about one and the same phenomenon. There is no “hard problem” here.

Owen Flanagan uses this argumentation in *Consciousness Reconsidered* (1992). Thus he ends in a position almost opposite of McGinn: where McGinn considers the gap between mind and brain as so wide that it could never be bridged (and that bridging in itself is senseless), Flanagan seems to consider it – in principle – as “hard” and “simple” as any other scientific problem. Here, however, I find it relevant to point out that the case of H₂O cannot be compared directly to the mind-brain problem: if one had a good enough microscope to study water, what we call H₂O becomes visible. Obviously, the mind will not appear as visible to a third-person observer by looking at the brain under a microscope. Furthermore, the mind is associated with a specific “first person perspective”, which can be said to make it much harder to explain by pointing to physical phenomena such as brains. Analogies to other scientific discoveries do not seem to hold water – so to speak – since they do not appreciate what makes the mind-brain problem different. Yet the critique raised by Flanagan (1992) – that there is no need to expect science to make us able to completely understand something by drawing inferences from another phenomenon – does not fall for this reason.

Natural science can in this context be conceived of as a methodology, with which the covariation of different phenomena can be observed. Based upon correlations, it can be argued that there is a connection between the two correlating phenomena, but it offers no understandings or explanations in itself. So, it can be argued that McGinn is actually just pointing out that science is not able to do something that it was never designed to do in the first place. If this is the case, it could be argued that the problems outlined by McGinn are not problems in the sense that they show a scientific account of consciousness to be impossible. Of course, it would still be just as unclear *how* it would be possible. This would in that case be dependent on our conception of what science is supposed to account for.

Methodological concerns

Even if one should conclude from the above that no principal argument succeeds in disqualifying consciousness as a scientific object, a closer analysis might reveal that there are methodological arguments that do.

How should one go about the methodological problems arising from the fact that conscious states are subjective and observable to the one having the experiences only? Indeed, all we observe when studying consciousness objectively are the reports given by subjects, not the conscious states themselves. So, are subjects' reports about conscious states actually about their conscious states – and: are such reports reliable indications of what is going on in experience?

Reports about consciousness. 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. So, this claimed difference is a difference of introspecting versus not introspecting one's conscious states. One could on this basis argue that a scientific approach to consciousness is based on the use of introspection, even though experimental psychology gave up on the introspectionism of Wundt and Titchener long ago.

Historically, however, the debate over introspection goes much further back than the introspectionists in the early period of experimental psychology. The earliest account of introspection – although the term "introspection" is not employed – may be found in Augustine's "De Trinitate" in which he states that "the mind should reflect upon itself" (1955, p. 80)⁴. Augustine suggests that a study of the mind should be based upon a technique of bracketing out the external world in order to perceive internal mental events or contents. The analogy used is one of inner perception or inner observation – one observes one's mental states as they are displayed. Augustine believed that such observations were flawless and always correct.

This latter claim was met by Thomas Aquinas who did not believe in a construction of a science nor a philosophy on the basis of introspection. He believed that introspection consisted in the mere sorting and understanding of ordinary external perception. This seems very much in line with many modern thinkers like Fred Dretske (1995), yet in the first half of the seventeenth century, introspection was taken very seriously. The works of René Descartes, also, should be taken as a product of this historical interest. In 1637, in his *Discourse on the Method of Rightly Directing One's Reason and of Seeking Truth in the Sciences*, he formulates his "foundationalist epistemology" by suggesting that one should

⁴ This, at least, I consider a "safe bet". Lyons (1986) has speculated that introspection may be traced all the way back to Aristotle, while e.g. Moustgaard (1990), more cautiously, "blames" John Locke's notion of an internal sense (Locke, 1690) as being the historically oldest mentioning of introspection.

only accept as true what presents itself to one's mind. In this way, the observation of the objects in one's experience was suggested as the foundation of science. Even Descartes' "archenemy", Thomas Hobbes, believed that a science of the mind should be based on self-reflection.

Although the term "introspection" came about in the latter part of the seventeenth century, the thinkers of the time did not successfully define the term, neither did they manage to properly disentangle concepts like "inner sense", "reflection", and "self knowledge" from introspection (Lyons, 1986). For instance, when John Locke speaks of our having ideas of reflection or when David Hume speaks of direct knowledge of our minds, it is not convincingly clear that they are addressing the topic of introspection – as the term later came to be known in experimental psychology.

Since the idea of introspecting conscious states goes much further back than Wundt's laboratory in Leipzig, I think one should not confuse introspection as such with classical introspectionism. Based on my choice of a definition of consciousness, I will stipulate the following definition of introspection:

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.

This definition, congruent with the definition of conscious states, has certain consequences:

1. Introspective states are always about conscious states. That is, one cannot be introspecting without being conscious in that being introspective involves an attending to consciousness.
2. Introspective states are always mental states but *not* conscious states as such. That is, there is no specific experiential quality that is introspective. This is believed by a number of theorists in favour of the earlier definition of consciousness⁵. However, introspection could be conceived of as a mental attitude to experiences, which may change the first-order experiences.

Criteria for the validity of reports. In the modern literature, there seems to be two lines of argument against the validity of introspective reports, although philosophers of the mind have not come much further in defining the term. One such argument is that we cannot deduce from behaviour (hence also reports) to conscious experience with absolute certainty. David Chalmers suggests with the alleged "zombie argument" that any observable human activity seems

⁵ According to the "higher order thought theorists", a, say, conscious perceptual state is dependent upon a subject having an in itself unconscious thought directed towards his perception. Introspection, then, involves an unconscious thought with the first thought and perceptual state as its content. It is important for the consistency of higher order theory that the thought is unconscious in that it would otherwise lead to infinite regress: The conscious thought must be dependent on yet another conscious thought that has it as its content, which again would be dependent on a conscious thought etc.

explainable in purely “functional” terms (Chalmers, 1996). This is not a problem for the behavioural sciences only, but just as much for introspectionism, in that verbal (or other) reports clearly also count as behaviour. The other objection is some variety of scepticism towards the use of subjective reports, normally with reference to memory being imperfect, being influenced by private associations that cannot be generalised and the like (Adams, 2000).

Obviously, a scientist would wish for a relation between consciousness and report as strong as possible, so that he could be certain that he was in fact studying consciousness when collecting subjective reports. He would wish for it to be possible to realise the following criterion:

1) For a public report to work as a publicly observable indication of consciousness, it should not be possible to observe the report under any other circumstances than when consciousness/the specific conscious state is present.

Accordingly, it could be argued that we can accept something publicly observable as a trustworthy indication of something that is subjectively observable only if the first is linked to the second by necessity. The problem is, of course, given the arguments stated above, if anything could actually live up to such a criterion.

Given the logical possibility of the existence of zombies proposed by Chalmers, and given that subjects are capable of lying, or just simply mistaking, nothing can meet the criterion except in the case of *my own* consciousness. Daniel Dennett is one of the very few people arguing, with an appeal to introspection, that “there is nothing there but functions” (1996). One can look to see if one finds cognitive functions or conscious experiences when introspecting, but it is very hard to lead a serious discussion about such issues and reply in a more sophisticated manner than, basically, “oh, yes there is” (as David Chalmers in fact does in 1997).

The case of my own consciousness is however far from good enough since it is not publicly observable. This might be used as an argument to state that a science of consciousness is an impossible mission for methodological reasons. However, it is also possible to argue that the criterion is stricter than necessary, and even stricter than what is often the case in other disciplines that we normally do not hesitate to label scientific. It is in fact very rare indeed that scientists claim to be dealing with “absolute certainty” in this manner; on the contrary, in the spirit of Karl Popper the critical rationalists, even so-called “hard sciences” such as physics or chemistry are only claiming to be testing beliefs. For this reason, one might alternatively suggest this criterion:

2) For a public report to work as a publicly observable indication of consciousness, the presence of consciousness should be the – intuitively – best possible explanation of why the report is given. The burden of evidence would then be with those who claim that this is not the case.

This view is of course not free of problems because something might be intuitively given to one person, but not to the next, and “givenness” is highly dependent on both social context and personal background. However, every science must at some level deal with intuitive givenness, in that even

scientists must evaluate arguments based on past experiences in the field, standard views of other scientists, etc., which eventually exerts an influence on what is “given” and what is not. It is quite common in science that one must argue more convincingly when stating something controversial, and one is allowed more *a priori* assumptions when stating the obvious. For instance, a brain scientist is definitely not expected to carry on discussions about the physical existence of the brain, wondering if objects are merely the result of his thinking about them, before he is allowed to conduct his research. So, in his field, the physical existence of the brain can easily be taken for granted. With reference to Sellars, as mentioned above, this is a matter of one’s choice of “unexplained explainer”, based upon which one’s explanation proceeds. In the same way, criterion 2 suggests taking the relation between reports about consciousness and consciousness itself for granted *a priori*. The advantage of this position as stated above would be that one does not have to take any *specific* understanding of the relation as *a priori* – only that reports about conscious states contain the best possible third-person data that can be collected about consciousness. That is, given that one accepts that conscious states are subjective states given to the first person, it follows that, when a subject says “right now I am experiencing this or that”, we have no means to falsify this utterance. That is, by reference to brain activations, behaviour, or cognitive performance in general, we can only second-guess what a subject experiences.

A different way of going against criterion 2 is to consider conscious experiences not only as subjective but also as *private*. A conscious state is subjective in the sense mentioned earlier: one subject only experiences it. However, saying that a conscious state is private, is to say that it cannot be communicated inter-subjectively. As has been argued, recently by Praetorius (2000), one cannot make sense of first person knowledge as private and beyond communication. Discussions of whether our descriptions of experiences refer to the same “things” as others refer to when they use the same terms, rest on the presupposition that experiences are things that we may have knowledge of and that we can describe. Every description of our experiences presupposes that as users of a language we know the correct applications of such descriptions. That is not to say that subjects cannot be mistaken or lying about their mental states, but that it is implied when accepting that, say, “red” names an experience that can be had by subjects that this term has a correct application and that it thus can be used to inform others about one’s experiential states. That is, the subjective character of experiences is not sufficient to claim that experiences are private or closed from descriptions that can be shared with others.

So, it is my position that reports about conscious experiences, with reasonable methodological requirements, can be considered reliable data.

The reader may now protest that this conclusion is very common sense and that any serious scientist, of course, in some sense would rely on subjective reports when studying consciousness.

However, this is far from the case. Many highly regarded theories in the field rest on very different assumptions. Take for instance, the following quotations:

Our basic idea is that consciousness depends crucially on some form of rather short-term memory and also on some form of serial attentional mechanism. This attentional mechanism helps sets of the relevant neurons to fire in a coherent semi-oscillatory way, probably at a frequency in the 40-70 Hz range, so that a temporary global unity is imposed on neurons in many different parts of the brain. (Crick & Koch, 1990, p. 263)

With this assumption, Crick & Koch actually go on to claim that one needs not use subjective reports when studying consciousness, but should instead look for neural networks in the brain, firing in a "semi-oscillatory way" and then deduce that "this subject or animal is now conscious of this or that".

Such suggestions are, as it shows, based on assumptions that have no empirical support. One could say that they would follow the logic of the syllogism below when doing research:

1. We believe that the conscious state Q is related to the publicly observable phenomenon X.
2. Organism Y (in state Z) exhibits behaviour X.
3. Ergo, organism Y (in state Z) entertains the conscious state Q.

However, the syllogism is based on another with the following structure:

1. We believe organism Y (in state Z) to entertain the conscious state Q.
2. Organism Y (in state Z) exhibits behaviour X.
3. Ergo, the conscious state Q is related to the publicly observable phenomenon X.

This latter syllogism is, of course, not very strong in that it relies on the truth-value of the belief in premise 1. Furthermore, the two syllogisms together form a circular argument where the assumption that Y is related to X is based on the assumption that Y in Z has Q, which is exactly what a scientific investigation in this framework aims to show.

However, one could now argue that this is no worse than what is being done when subjective reports are used to collect data. In fact, if one would put in "report" as "X" above, the same sort of logic is revealed. Based on this line of thinking, the threat from a "zombie-like argument" is very close.

There are a number of methodological advances when using subjective reports instead of "something else" like brain waves or any other functional criterion. For one thing, subjective reports can be validated through interviews (asking the subjects if they are sure that they only gave the reports under the correct circumstances, if it is really true that they do give reports every time they experienced this or that etc.), whereas no such introspective technique can be employed when using a measure beyond the subject's control (such as oscillatory firing of neurons). But apart from such technicalities, one seems to be forced to admit that it is in fact

possible to cast doubts on the fundamental methodology of a science of consciousness with regards to the truth value of reports. On the other hand, however, in doing so, one would place oneself in an extreme position arguing that the subjects are zombies, robots or liars. Such a position, it would seem, would not only have consequences for laboratory subjects; it would follow that people at large are zombies, robots or liars. Such a position may not be desirable, logically consistent or not.

It is often said that experiencing is not some fundamental given about which we cannot be mistaken (e.g. Lyons, 1986). This, however, does not harmonise well with an understanding of consciousness as a state, one is *directly* aware of being in. The solution to this problem, I believe, lies in a disentangling of our conscious states and our knowledge about our conscious states.

Knowledge is sometimes predicated of experiencing, sometimes of reports about experience, and sometimes of true claims based on experience. Thus, we can say we have known pain, meaning that we have experienced pain. Or we report to others that we are experiencing pain, and we characterise the basis of our report as privileged knowledge. Or we claim that someone else is in pain, saying we know it from their behaviour. The differences between these kinds of knowledge are crucial to develop a valid methodology for a science of consciousness.

In the case of experiencing pain, the subject cannot be mistaken about the experience itself, because it is simply as it seems. No claim is yet made about it, which can be judged as right or wrong. Therefore, to say that an act of experiencing is knowledge is simply to equate the two words. Before the subject can be mistaken, he or she must claim something, which the subject does only by way of some sort of report. An expression, such as a grimace when in pain, is not sufficient, in that this, also, is no claim. Only in the case of the report lies a claim that "I am in pain". This claim is not falsifiable as such for an external scientist, due to our not having access to the contents of other minds, yet it may be true or false nonetheless.

So, while experiencing is immediately or directly given to the subject, some or other cognitive processes underlie giving a claim about the experience.

It is crucial to insist that the subjective reports that are taken as third person indications of conscious states are in fact reports that explicitly refer to the experience of the subjects. If a scientist is, say, to ask a subject to discriminate between two colours, he does not introduce an introspective task (Merikle & Rheingold, 1992). Not least since the finding by Marcel (1983) that unconscious perception could occur even in the absence of the ability to discriminate, it has been widely accepted that perception can occur unconsciously. Yet, it has survived as a dominant paradigm to measure the presence of visual conscious experiences to see whether subjects can perform perceptual functions.

If the scientist instead asks the subject if he has *seen* a difference between two colours, it is unclear what kind of task the subject performs. The question "have you seen a difference between the two colours?" can be interpreted as 1) was there a difference (guess)? or 2) did you experience a

difference? Or, very likely, a subject may use the two interpretations interchangeably.

Therefore, ideally, the scientist should use very explicit instructions: "did you experience a difference between the two colours?" and train his subjects to understand the difference between guessing and experiencing.

In experimental psychology, there has been a long-lasting debate over whether one should use trained subjects. It may be problematic to extrapolate the data collected in an experiment to the rest of the population when the experimental subjects have undergone a procedure to teach them to respond differently than what they otherwise may have done. That is, it sounds suspicious, *prima facie*, to train the subjects' performance in order to learn about the performance of people in general. Or, in other words, the object one wishes to study might change in unforeseen ways. Furthermore, it could be argued that training not only changes the kinds of responses the subjects use to express their conscious states – training may change the experience itself (Adams, 2000).

The consequences of *not* training the subjects is however to accept a high uncertainty concerning the knowledge about their experiences. That is, the reports of the subjects would be much harder to interpret, and there would be no objective way to control for this⁶. One could also argue that it is *not* so that the subjects were somehow "uncontaminated" before the training: each of them would give reports about their experiences based upon some pre-existing beliefs (about the concept of consciousness, about how to behave in experimental settings, etc.). If this is the case, the training does nothing but control for this effect. Finally, one could argue that even if training would change the experiences of the subjects, not only the behaviour, it would have no importance as long as we are looking for the way in which consciousness is generated. In a sense, the critique that training changes the experiences is only a real critique insofar as one is trying to study experiences under specific circumstances: in some kind of "natural setting", whatever that would mean. If this is the case, the critique should not just be raised against training, but against laboratory settings as such. However, the present study does not involve such a specification, and as such, the critique seems less relevant.

This view of how to operationalise consciousness supplies specific guidelines when studying consciousness in laboratory settings. With use of subjective reports and explicit reference to the conscious states of the subject, one can obtain scientifically valid data to compare to, say, neuroscientific or behavioural data. However, it has consequences for qualitative research also, as it is a general theoretical framework with which one can have access to the thoughts and feelings of other people, and at which level of certainty one can do so. As such, it may even be said to give a theoretical background to our every-day conceptions of other people.

⁶ At least in the framework provided by "criterion 2".

The neural correlate of consciousness

The neural correlate of consciousness (NCC) is the single line of consciousness research in which most efforts are made; first and foremost because most current positions in the mind-body debate are of a materialistic persuasion, and, according to such positions, finding the NCC is of essential explanatory value.

It is however not clear what it means that the relation between a conscious state and a brain state form a correlation. However, to begin with, I find it reasonable to look into how the idea of an NCC is grasped in consciousness research in general.

The simplest working definition of an NCC would be something like:

A neural system is an NCC if it correlates directly with states of consciousness.

But to understand even this, we need to understand a) the notion of "direct correlations", and b) which notion of "states of consciousness" we seek to operationalise.

The notion of direct correlations of consciousness. What do we expect of a neural system to let it qualify as an NCC?

One possibility would be to say that such a neural system should be both necessary and sufficient for conscious states, so that there will never be a conscious state without the system, and there would be no need to include other neural systems as NCC's. I will agree with David Chalmers (2000) here, that this notion might prove too strong. First of all, there might be more than one neural state corresponding to each mental state, as it is claimed by functionalism, so that one neural system is sufficient for being in a particular conscious state, while another neural system also suffices, even though the two do not have to be correlated all the time. If neural system 1 can be active, giving rise to a particular conscious state while system 2 is not active, and it also were true that if 2 was active and 1 inactive, the same conscious state would be present, neither of the two could be considered necessary for consciousness, though possibly sufficient.

So, we could just say that an NCC must be sufficient for consciousness. However, this would reduce the explanatory value of hunting NCC's immensely, since there probably cannot be found a particular neural state that is not active when in some conscious state, at one time or another. For one thing, this would hardly live up to common standards of neuroscience, where a neural correlate must be more closely related to the mental phenomenon in question, and the actual NCC would in this case be something like the entire nervous system (all the brain, definitely). The trouble with accepting neural systems unnecessary for consciousness as parts of an NCC is that it leads to accepting irrelevant features as parts of an explanation.

One answer to this could be that one should seek a *minimal* sufficient system (Chalmers, 2000). A simple strategy to achieve the minimal sufficient system is to somehow ensure that no part of the neural system in itself is

sufficient for consciousness. In this way, features irrelevant for the explanation could be excluded. The “end product” with an exclusion strategy would still be different from a necessary neural system, to which it would be an absolute demand that the conscious state could not occur without it (due to the possibility of several systems corresponding to the same conscious states).

Some neural systems may of course be necessary for consciousness without actually being necessary parts of an explanation, like for instance the brain stem, without which there would hardly be any mental activity at all. Yet, to use a radio metaphor, one would not consider the on/off switch a convincing explanation of the specific functions of the radio – for one thing, the radio and other electrical instruments all have switches, and they are turned on no matter which specific function the instrument performs. For a “true” NCC, it must be the case that whenever it is active, the subject is in a corresponding conscious state (though the opposite may not be true), which cannot be said about brain areas that mainly just activate other areas.

The notion of states of consciousness. A “state of consciousness”, as I have used it in the paragraph above, can have several different meanings. In most cases, a state of consciousness is either interpreted as the specific content of a conscious experience, or as “consciousness as such”. In the first understanding, an NCC is a neural system minimally sufficient for having *this specific* experience – for representing something (and not something else) as an experience to the subject. This obviously includes anything from seeing the colour red to a thought process or an unlocalisable and unrecognisable feeling. Research on content-NCC’s must primarily deal with neural differences (say, locations in the brain) when having different experiences, correlating some sort of introspective reports and measures of the brain. The over-all question would be something like: what (neural) properties correlate with being conscious of X that are different from being conscious of Y or Z? This seems to be the strategy most widely used in current consciousness research, which in fact is not so different from more traditional cognitive science. The major difference would however still be that the NCC would not represent the functional aspect of the mental phenomenon, but only its phenomenal properties. So a definition of a neural correlate of content-consciousness would be:

A neural system is a neural correlate of content of consciousness if it correlates directly with a state of having one particular conscious experience and if it does not correlate with every other conscious or any unconscious state.

NCC’s directed at consciousness as such are a different matter, where the phenomenal property in itself is

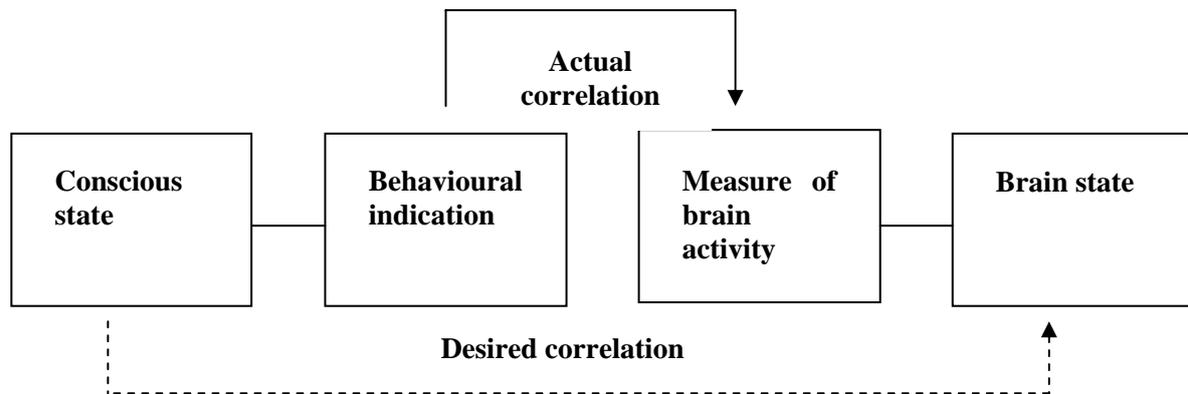
the object of research, and where the specific contents of consciousness are thought of as special cases of this property. Here, the research would be oriented towards what properties that correlate with being conscious of X or Y that are different from not being conscious of X or Y. With the intuitive notion that consciousness is a specific and distinct property, it is a reasonable hypothesis that there would be distinct NCC’s that are non-specific in terms of content of consciousness. Furthermore, it seems plausible that this kind of NCC will be something like a type of activity and not a specific brain area. There are several reasons for this: first of all, there has never been observed a brain region which is always activated in conscious states or to which all information in the brain is sent (Dennett & Kinsbourne, 1992). Second of all, one practical implication of such a “consciousness module” would be that it would be possible to get a lesion that would make you completely unconscious, but still leave all of your mental functions intact and normal. This would of course never be detected, even if it could happen, since such a functional zombie would be like any “normal” person in terms of behaviour, and thus it does not function as a convincing argument in itself. A stronger argument would be that it seems like a misunderstood description of consciousness that it should be a result of connection between a “consciousness area” and a “dead area”. Besides, with such an organisation of the brain, consciousness could be achieved by electrically stimulating the “consciousness module”, and thus having content-less consciousness. Taken together, I find that it is highly plausible that a neural correlate of consciousness as such would be something like levels of activity, so that any system that is in that particular activation state gives rise to a conscious state with a particular content. This hypothetical assumption is shared by a number of researchers, e.g. Libet (1985) and Edelman & Tononi (2000).

A concept of a neural correlate of consciousness as such, can be defined as follows.

A neural system is a neural correlate of consciousness as such if it correlates directly with any possible conscious state and if it does not correlate with any unconscious state.

A “linking problem” for NCC research

However, the notion of “direct correlations” is brought into doubt by the problems that the notion of a linking principle points out. It can be illustrated like this:



By “conscious state” I mean a phenomenal state as it is to the subject. The “behavioural indication” refers to anything observable that is believed to be closely related to consciousness. This might be verbal reports about conscious states, button pressings by experimental subjects, certain behavioural patterns by animals etc. “Measure of brain activity” refers to the data one gets from PET, EEG, fMRI or any such technique, and “brain state” is then, of course, the actual brain state “an sich”.

The relation between the behavioural indication and measures of the brain can be examined scientifically by looking for statistical correlates, but this is not identical to an examination of the proposed relation between conscious and brain states *per se*. The relation between the conscious state and the brain state is the NCC. In this sense, it seems fair to say that the NCC is more of a theoretical abstraction than something that can actually be measured directly.

Now, one could ask if it is not the case in many branches of science that the object of study cannot be measured directly. This seems indeed to be the case. Take the example of gravity. Gravity is not directly observed – it is an inference made to explain other observations. Or the example of atom physics. Protons and electrons are not observed directly, yet we can obtain quite strong theories of their existence based upon other observations that are thought to be explainable by inferring the existence of these little particles. The same kind of example could also come from the social sciences, where social dynamics are not observed directly, but are inferences as well. However, none of these examples are identical to the problem of consciousness. Consciousness is *not* an inference made to explain behaviour; it is directly observable in itself. That is, you do not need to make any kinds of inferences to determine if you are conscious. In the same way, the behaviour that is believed to indicate the presence of a conscious state is directly observable.

The model illustrates that when philosophers and scientists talk about NCC’s, they are *in fact* talking about “NCRC’s” – Neural Correlates of Reports on Consciousness. This means that the idea of a “clean” NCC amounts to something more theoretical, namely an “indirect” correlation

between conscious and neural states, resting on the strength of the connection between the experience and the report, accessible through a third person perspective.

Conclusion

So far, I have argued the following:

- One can have a non-circular definition of consciousness that implies a subject who is directly aware of his or her mental states.
- Arguments so far presented saying that a science of consciousness is impossible for reasons of principle all misconstrue the problems in the operationalisation of conscious states.
- Certain methodological problems exist due to the subjectivity of conscious states. The only way out, seemingly, is to accept a relation between reports about conscious states and the conscious states themselves as an *a priori*.
- Still, in order to make use of this relation, one must consider the factor of introspection as well as the general truth-value of subjective reports.
- “Neural correlates of consciousness” should be investigated as “minimally sufficient” neural systems.
- Neural correlates of consciousness resemble theoretical abstractions more than direct observations until we have a clear understanding of the relationship between conscious states and the reports about them, and neural states and the measures of them.

In the course of my argumentation, I hope to have shown that it is in fact possible to give a definition of consciousness that lives up to formal criteria in analytical philosophy and, simultaneously, relates to our every-day understanding of the term. The operationalisation leaves us with an account of subjective reports and their relations to conscious experiences that are of use in experimental settings

as well as in other, less construed interactions with other people. Clearly, the operationalisation here is not complete, and one could justly criticise me for not having included all aspects of a full research programme. For instance, the operationalisation and background theory for making laboratory work on general psychological concepts should also be reflected in the analysis of data. However, I hope to have made just a few steps forwards in the attempt to make progress in the establishing of this theoretical framework. For a more elaborated version of these ideas, please consult Overgaard, 2003.

It has been my argument that with the use of a terminology that is not dependent upon an acceptance of the theoretical frameworks of the cognitive sciences, and that is open to concepts used in general psychology, so-called general psychologists would be able to make relevant operationalisations and empirical testings of their theories. Furthermore, the fact that the framework here presented does not imply reductionism, and that it links together with concepts and understandings that can be recognised in our every-day existence, will make a cognitive science along these lines inspire us to think of ourselves as biological yet conscious, irreducible creatures: humans, so to say.⁷

References

- Adams, W. (2000): Introspectionism reconsidered, paper presented at *Toward a Science of Consciousness*, April 10-15, Tucson, USA
- Augustine (1955): The Trinity, in Augustine: *Later Works*, vol. 8, ed. and trans. J. Burnaby, SCM Press
- Block, N. (1995): On a confusion about a function of consciousness, *Behavioral & Brain Sciences*, 18, 2, 227-287
- Chalmers, D.J. (1996): *The Conscious Mind*, Oxford University Press
- Chalmers, D.J. (1997): Moving forward on the problem of consciousness, *Journal of Consciousness Studies*, 4, 3-46
- Chalmers, D.J. (2000): What is a neural correlate of consciousness? in: T.Metzinger (ed.): *Neural Correlates of Consciousness*, MIT Press
- Chrisley, R. (2001): A view from anywhere: prospects for an objective understanding of consciousness, in P. Pyllkkänen & T.Vadén (eds.): *Dimensions of Conscious Experience*, John Benjamin Publishers
- Chrisley, R. (2002): Is there a seemings-is distinction for seemings? paper presented at the *First Meeting of the Nordic Network for Consciousness Studies*, May 3-4, Skövde, Sweden
- Crick, F. & Koch, C. (1990): Toward a neurobiological theory of consciousness, *Seminars in the Neurosciences*, 2, 263-275
- Dennett, D.C. (1996): Facing backwards on the problem of consciousness, *Journal of Consciousness Studies*, 3, 1, 4-6
- Dennett, D.C. & Kinsbourne, M. (1992): Time and the observer: There where and when of consciousness in the brain, *Behavioral & Brain Sciences*, 15, 183-247
- Descartes, R. (1637/1954): *Discourse on the Method of Rightly Directing One's Reason and of Seeking Truth in the Sciences*, in: R.Descartes: *Philosophical Writings*, ed. and trans. E. Anscombe and P.T. Geach, Nelson
- Dretske, F. (1995): *Naturalizing the Mind*, MIT Press
- Edelman, G.M. & Tononi, G. (2000): Reentry and the dynamic core: neural correlates of conscious experience, in T.Metzinger (ed.): *Neural Correlates of Consciousness*, MIT Press
- Flanagan, O. (1992): *Consciousness Reconsidered*, MIT Press
- Jack, A. & Shallice, T. (2001): Introspective physicalism as an approach to the science of consciousness, Cognition special issue edited by S. Dohaene: *The Cognitive Neuroscience of Consciousness*, MIT Press
- Jackson, F. (1986): What Mary didn't know, *Journal of Philosophy*, 83, 291-295
- James, W. (1890): *The Principles of Psychology*, Dover
- Krakow, I. (2002): *Why the Mind-Body Problem cannot be solved*, University Press of America
- Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action, *Behavioral and Brain Sciences*, 8, 529-66.
- Locke, J. (1690/1959): *An Essay Concerning Human Understanding*, Dover
- Lyons, W. (1986): *The Disappearance of Introspection*, MIT Press
- Marcel, A.J. (1983): Conscious and unconscious perception: experiments on visual masking and word recognition, *Cognitive Psychology*, 15, 197-237
- Marcel, A.J. (1993). Slippage in the unity of consciousness. In: G.R.Bock & J.Marsh, (Eds.): *Experimental and Theoretical Studies of Consciousness*. John Wiley & Sons, 168-186.
- McGinn, C (1989) Can we solve the mind-body problem? *Mind*, 98, 349-366
- McGinn, C. (1991) *The Problem of Consciousness: Essays Toward a Resolution*, Blackwell
- Merikle, P.M. & Rheingold, E.M. (1992): Measuring unconscious perceptual processes, in R.F. Bornstein & T.S. Pittman (eds.): *Perception without Awareness*, Guilford Press
- Nagel, T. (1974): What is it like to be a bat? *Philosophical Review* *The Philosophical Review*, LXXXIII, 435-51
- Nagel, T. (1986): *The View from Nowhere*, Oxford University Press
- Moustgaard, I.K. (1990): *Psychological Observation and Description*, Sigma Forlag
- Overgaard, M. (2003): *Theoretical and Experimental studies of Consciousness*, Ph.D.-thesis, University of Aarhus
- Overgaard, M., Kauffmann, O. & Ramsøy, T.Z. (2001): Consciousness and introspection, *Consciousness Research Abstract: Journal of Consciousness Studies*, 182
- Praetorius, N. (2000): *Principles of Cognition, Language and Action*, Kluwer Academic Publishers
- Revonsuo, A. (2001): Can functional brain imaging discover consciousness in the brain? *Journal of Consciousness Studies*, 8, 3,
- Searle, J. (1992): *The Rediscovery of the Mind*, MIT Press
- Sellars, W. (1969): The language of theories, *Philosophy and Phenomenological Research*, 29, 1969
- Varela, F. (1999): The specious present: a neurophenomenology of time consciousness, in: J.Petitot et al. (eds.): *Naturalizing Phenomenology*, Stanford University Press.

⁷ I wish to thank Allan Jones, dept. of Psychology, University of Aarhus for proofreading my English.