

# Cognitive Approaches to Phenomenal Consciousness

Peter Mandik

## 1 Introduction: Cognition and cognitive approaches to phenomenal consciousness

To my mind, the most promising approaches to understanding phenomenal consciousness are what I'll call *cognitive approaches*, the most notable exemplars of which are the theories of consciousness articulated by Rosenthal (2005, 2011) and Dennett (1991, 2005). The aim of the present contribution is to review the core similarities and differences of these exemplars, as well as to outline the main strengths and remaining challenges to this general sort of approach.

Cognitive approaches to phenomenal consciousness give explanatory pride of place to *cognitive states* – states such as judgements, thoughts and beliefs – in explanations of *phenomenally conscious states*, prototypical instances of which include the visual experience of seeing the vivid red of a ripe tomato. Such an approach may initially seem puzzling, given a seemingly sharp contrast between cognition and experience. This seeming contrast can be illustrated in a manner due to Sellars (1997). Pre-theoretically, there seems to be an obvious distinction between, on the one hand, (1) *thinking* that the banana in my lunchbox is yellow without at the same time either *seeing* the banana as yellow or anything at that time *looking* yellow to me, and, on the other hand, either (2a) *seeing* the banana as yellow when it is in fact yellow, (2b) it *looking* to me that the banana is yellow when it is in fact some other colour, or (2c) it *looking* to me that there is a yellow banana even in the absence of either bananas or yellow things. The key allegedly sharp contrast here is that between thinking and seeing. An opponent of the cognitive approach to phenomenal consciousness holds that the sorts of explanatory resources adequate for an account of (1) cannot without supplement be used to give an adequate

account of the sorts of states exemplified by (2a–c). One way of opposing the cognitive approach would be to hold that what distinguishes states of type (2) from states of type (1) is that type-2 states but not type-1 states essentially involve the presence of a sensation, for example, a visual sensation of yellow, and further, that sensations cannot be explained utilizing only the resources minimally adequate for explaining type (1) states. On some elaborations of this sensation account, sensations are states distinguished by the possession of a certain kind of property, so-called phenomenal properties or qualia (plural), in the present case, a yellow quale (singular). In contrast, the proponent of the cognitive approach holds that the explanation of states of type (2) being phenomenally conscious can be adequately spelled out in terms of the minimal resources needed to explain states of type (1). Distinctive of what I'm presently calling *cognitive approaches to phenomenal consciousness* is optimism about what I'll call a 'reductive' explanation of phenomenal consciousness in terms of cognition – a non-circular explanation of phenomenal consciousness in terms of cognition, an explanation that neither explicitly nor tacitly presupposes that cognition must be explained by reference to phenomenal consciousness. (See Byrne's (1997) similar use of 'reductive' (pp. 103–4).)

In the present work, I will review two prime examples of cognitive approaches to phenomenal consciousness. The first is David Rosenthal's higher-order thought theory of consciousness (Rosenthal 2005, 2011). The second is Daniel Dennett's fame in the brain theory of consciousness (2005), previously known as his multiple-drafts theory (Dennett 1991).

One of the issues I will spend some time on is that of whether these approaches to phenomenal consciousness are best viewed as explaining it or instead explaining it away or maybe even bearing some altogether different relation to it (e.g. deliberately ignoring it). The question here is what the exemplars of cognitive approaches are best seen as offering explanations of, and to what degree they thereby are in the ballpark of satisfying requests of those calling for an explanation of so-called phenomenal consciousness.

## 2 David Rosenthal's higher-order thought theory of consciousness

At the heart of Rosenthal's higher-order thought theory is the idea of a higher-order thought, a thought that one has about one of one's own mental states. (In contrast, a first-order thought is a thought not about one's own mental

states.) The gist of Rosenthal's theory is the view that what it means for one of your mental states to be conscious is that you have a higher-order thought about that state. Often, in presenting the theory, Rosenthal makes note of several distinct uses of the word 'conscious', and notes that only one of them describes the explanatory target of Rosenthal's theory. This is the use of the word 'conscious' where we apply it as an adjective describing mental states as being conscious (as opposed to unconscious) mental states. This way of using the word conscious picks out what Rosenthal calls *state consciousness*. The aim of Rosenthal's theory is to explain what state consciousness consists in. Another adjectival use of the word conscious applies to entire people or animals as when we say that a recently awakened child is conscious or a boxer who is knocked out is unconscious. This latter use of the word conscious picks out what Rosenthal calls *creature consciousness* and it is less central to Rosenthal's theory than state consciousness. A conscious creature is a creature who is awake and responsive to stimuli. However, the conscious creature can have multiple mental states at a single time which differ in that some of the states are conscious states while others are unconscious states. What this difference consists in is the central thing that Rosenthal is seeking to supply an explanation of. A third use of the word conscious picks out what Rosenthal calls transitive consciousness. Examples of transitive consciousness include those cases in which we would describe someone as being *conscious of* something.

Transitive consciousness plays a crucial role in Rosenthal's explanation of state consciousness. Rosenthal endorses a link between transitive consciousness and state consciousness that he calls the *transitivity principle*. According to the transitivity principle, a person's mental state is conscious in virtue of that person being conscious of that state. The transitivity principle is supposed to capture a pre-theoretic view that people hold about state consciousness. One way in which Rosenthal expresses the transitivity principle is by stating that we would not regard a state as conscious if the person having the state was in no way conscious of being in that state.

Transitive consciousness comes in at least two varieties, only one of which is central to Rosenthal's theory. Generally speaking, when it comes to being conscious of things, there are several ways in which a person can be transitively conscious. One way is perceptual: I am conscious of my coffee mug right now in virtue of seeing my coffee mug. However, argues Rosenthal, another way in which we could be conscious of things is by thinking about them. In thinking about my cat even though I may not be currently perceiving the cat, I am nonetheless, in virtue of having that thought, thereby conscious of the cat.

Since there are perceptual exemplars of transitive consciousness, the theoretical possibility opens of consciousness of one's mental states as being mediated by some sort of perceptual relation, an 'inner sense'. Indeed, some thinkers have endorsed a higher-order perception account of the consciousness of one's own mental states. See, for example, Lycan (1996). Rosenthal argues against higher-order perception accounts of state consciousness on the grounds that (1) perception is always mediated by the presence of what he calls a mental quality and (2) his claim that our consciousness of our own mental states is not itself mediated by any comparable mental quality. This leads Rosenthal to his view that the way in which we are transitively conscious of our mental states that renders those states conscious is in virtue of having thoughts about those states namely, an explanation of consciousness in terms of higher-order thoughts.

One crucial feature to note about Rosenthal's theory of consciousness is the way in which it attempts to supply a non-circular explanation of consciousness. There may initially seem to be circularity in the theory since he is explaining a state's being conscious in a virtue of a person being conscious of that state. However, the crucial distinction between state consciousness and transitive consciousness helps to block against such circularity. It would of course be circular to explain state consciousness in terms of state consciousness and also circular to explain transitive consciousness in terms of transitive consciousness. But such circularities are not what Rosenthal offers. His non-circular explanation is an explanation of state consciousness, which is one thing, in terms of transitive consciousness, which is something else. Of course, in virtue of the state's being conscious, there must be some other state which is about the first state. However, on pain of circularity, that other state must be allowed to be a non-conscious state.

One way to make the non-circularity of the higher-order thought theory most apparent can be illustrated in terms of two states. The first state is a first-order state, for example a perception of a red apple. States of this type can occur unconsciously as in subliminal perception or they can occur consciously. When they occur consciously they do so in virtue of there being a higher-order thought about the first-order state. So, in virtue of having a higher-order thought that one is perceiving a red apple, one has a conscious first-order state, namely the conscious perception of a red apple. The higher-order thought need not itself be conscious. If it were conscious it would need to be conscious in virtue of there being some third-order state about it. But in the absence of any such third-order state, the second-order state is unconscious. The first-order state is conscious in

virtue of being targeted by the non-conscious second-order state. We have here a sketch of a reductive explanation of state consciousness since state consciousness is being explained in terms of things none of which are individually themselves conscious states.

Before closing the section there are two points that I want to make. The first concerns Rosenthal's explicitly stated explanatory target and its relation to so-called phenomenal consciousness. The second concerns the puzzling case of empty higher-order thoughts – higher-order thoughts that are about first-order states that themselves do not exist. Rosenthal's treatment of empty higher-order thoughts serves to illustrate the load that cognitive states are meant to bear in his explanatory project.

## 2.1 Phenomenal consciousness and Rosenthal's explanatory target

As Rosenthal makes quite clear, his central explanatory target is state consciousness. One might wonder, then, what pertinence Rosenthal's project might have for so-called *phenomenal consciousness*. Many authors writing on consciousness make phenomenal consciousness their central concern. For example, Chalmers' famous work promoting the hardness of the 'hard problem' of consciousness is explicitly cast as being about the difficulty in explaining *phenomenal* consciousness (1996). When Velmans and Schneider assembled their authors for their Blackwell Companion to consciousness (2006), they asked authors to be sure they addressed phenomenal consciousness (see my own contribution to that volume (Mandik 2006)).

In explicating what they mean by 'phenomenal consciousness' many, and perhaps most, authors working in contemporary philosophy of mind connect that phrase to a certain use of the phrase 'what it is like', especially as that phrased is utilized in Nagel's seminal 'What is it like to be a bat?' (1974). In Nagel's paper, he highlights a problem concerning knowing what it is like to be a bat. Not being bats ourselves, we have no introspective access to the mental lives of bats. From a third-person point of view, we note that their sensory systems are very different from those of humans, and we infer that it is very likely that what it is like to be a bat is likely to be very different from being a human. But how, if at all, can the objective third-person methods common in the physical sciences allow us to know what it is like to be a bat? A similar sort of problem concerning knowledge is highlighted in Jackson's (1982) famous thought experiment concerning Mary, a hypothetical colour-blind neuroscientist who knows all of the physical facts about what happens in human brains when humans see red,

but, since she has not herself seen red, it seems intuitive to many that Mary would not know what it is like to see red.

Many authors in the philosophy of mind connect the ‘what it is like’ phraseology central to explicating ‘phenomenal consciousness’ to the philosophical technical term qualia. (See, for example, Frankish (ms) and 2012). In brief, qualia are phenomenal properties, properties of mental states in virtue of which ‘there is something it’s like’ to be in them. (For an elaboration on worries concerning the technical terms ‘qualia’ and ‘phenomenal consciousness’ see Mandik (forthcoming).)

Nagel connects the idea of consciousness to the ‘what it is like’ phraseology by explicitly characterizing conscious states as states it is like something to be in. As Nagel puts it: ‘An organism has conscious mental states if and only if there is something that it is like to be that organism’ (Nagel 1974, p. 436).

Rosenthal explicitly addresses the connection between Nagelian what-it-is-like talk and his own project of explaining state consciousness. He writes (1997, p. 733) ‘What it is like for one to have a pain, in the relevant sense of that idiom, is simply what it is like for one to be conscious of having that pain.’ He also writes (2011: 433–34):

As many, myself included, use that phrase, there being something it’s like for one to be in a state is simply its seeming subjectively that one is in that state. ... And ... a HOT is sufficient for there to be something it’s like for one to be in the state the HOT describes, even if that state doesn’t occur.

We might summarize Rosenthal’s views here in the following manner: The ‘what it is like’ phraseology as most pertinent to discussions of consciousness picks out the way in which one’s own mental life seems to one, that is, the subjective appearance of one’s own mental life to oneself. Further, the way in which one’s own mental life appears is fully determined by certain thoughts one has. So, for instance, it’s seeming to me that I am in pain just is me having a certain thought to the effect that I am in pain. Whether I really am in pain is a separate matter, but we are here concerned with appearance, not reality, and the appearances in question are determined by the having of certain thoughts, regardless of whether the thoughts are accurate or not. The connection of what-it-is-like phraseology to state consciousness, then, is that in having the thoughts determinative of the subjective appearances picked out by ‘what it is like’ phraseology, one thereby has, in virtue of having those thoughts, states that are conscious. The conscious states are the states the thoughts are about. The conscious states are the states it appears to oneself that one is in.

## 2.2 Empty higher-order thoughts and the centrality of cognition in Rosenthal's theory

One interesting feature of Rosenthal's theory and the feature that especially **a** highlights the central role that cognition plays in explaining consciousness is a feature that we can call attention to by contemplating false higher-order thoughts. It is a general feature of thoughts that they can sometimes be false, and Rosenthal sees no reason to exclude consciousness-conferring higher-order thoughts from such a generalization. Rosenthal thus leaves open the possibility, for example, of a higher-order thought that one is having a perception of a red apple when in fact one is either having a perception of a green apple or having no perception at all.

The case of empty higher-order thoughts is somewhat puzzling, at least initially, as is Rosenthal's treatment of the case. However, some clarity can be gained by making explicit two kinds of readings of the HOT theory – a relational reading and a non-relational reading – and emphasizing which of the two Rosenthal intends, namely, the non-relational one.

One way of spelling out the issues at play here is to consider a three-way comparison between three highly similar people who all have the same higher-order thought but differ only in some respects that are extrinsic to their respective HOTs. Persons A, B and C, let us suppose, all have a HOT, the content of which would be expressible by 'I am currently having a visual experience of a red apple'. Let A's HOT be true: A is in fact currently having a visual experience of a red apple. Let both B's HOT and C's HOT be false. Let B's HOT be false because B is having a visual experience of a green apple. Let C's HOT be false because C is not in fact currently having any visual experiences. Let us call the case of C *the case of the empty HOT*, since the first-order state that the HOT is about doesn't even exist. Especially interesting here is a comparison between C, the empty case, and A and B, the cases in which the relevant HOTs are non-empty. Now consider this question: According to HOT theory, are the empty and non-empty cases alike in that each of the people described have a currently conscious visual state or, instead, does one of them, namely the empty case, C, *lack* a current conscious visual state? The different readings of the HOT theory, the relational reading and the non-relational reading, supply different answers to this question.

According to the relational reading, a person is in a conscious state if and only if they are in an actually existing first-order mental state and they are also in a higher-order state about that first state. The qualifier 'actually existing' is crucial in distinguishing the relational from the non-relational reading. According to

the non-relational reading, having a higher-order thought alone suffices for being in a conscious state. There need not, then, be an actually existing first-order state that the HOT is about. The state the HOT is about may, as when the HOT is empty, be a merely notional state.

On the relational reading of HOT theory, a state's being conscious is a relational matter. One and the same first-order state can be unconscious at one time and conscious at another, and what determines this change from unconscious to conscious is nothing intrinsic to the first-order state itself, but instead the change from not being related to a suitable HOT to being so related. For commenters who have adopted the relational reading, see Bruno (2005), Gennaro (2006, 2012) and Wilberg (2010). On the assumption that the instantiation of a two-place relation requires the existence of both of its relata (see Kriegel 2007, 2008 and Mandik 2009), the case of C, the person with an empty higher-order thought, seems to be a case in which the person lacks a conscious state. Although they have a HOT, they do not have any actually existing first-order state that the HOT is about, and so C would be in the puzzling circumstance of lacking a conscious state while seeming to themselves to have one. What's puzzling about this is that the way the mental lives of A, B and C each seem to them would seem to be the same – it seems to each of them, in virtue of having the same HOT, that each of them has a first-order state of perceiving a red apple. If phenomenal consciousness is a matter of how one's own mental life seems to one, then in reading HOT theory relationally, we see state consciousness and phenomenal consciousness coming apart. If A is phenomenally conscious, then so is C. But C lacks a conscious state. So C is phenomenally conscious without having a conscious state, since having a conscious state is here interpreted as having an actually existing first-order state that one is conscious of in virtue of having a HOT about it. (For a view that explicitly endorses separating state consciousness from phenomenal consciousness, see Brown (2015).)

The non-relational reading has the advantage of not separating phenomenal consciousness from state consciousness. However, it accomplishes this by allowing that when a person is truly describable as having a conscious state, the state in question may on some occasions be merely notional. But this is of a piece with interpreting consciousness as being a matter of appearances, a matter of how one's own mental life appears to one.

(Though it is worth noting, space does not permit discussion of a possible third reading, a disjunctive theory of state consciousness that combines resources from the relational and non-relational readings of HOT theory. In brief, such a disjunctive theory holds that a disjunctive condition suffices for one's being in a

conscious state – one is in a conscious state if *either* there is an actually existing first-order state and a higher-order thought bearing an ‘aboutness’ relation to it *or* there is simply a higher-order thought and no actually existing first-order state that it is about. Like the non-disjunctive relational reading, the disjunctive reading runs afoul of the sorts of problems with a so-called aboutness relation I spell out at length in Mandik (2009.)

### 3 Daniel Dennett’s Fame In the Brain Theory of Consciousness

#### 3.1 From Rosenthal to Dennett: Dennett’s Zimbo argument

In Dennett’s seminal exposition of his theory of consciousness, his 1991 book *Consciousness Explained*, he temporarily borrows Rosenthal’s HOT theory to articulate his own now-famous ‘zimbo’ argument against zombies. In philosophy of mind, the term ‘zombie’ is used to denote not some undead flesh-eater from horror cinema, but instead a being that is similar to a normal human in some major respect (e.g. behaviourally, functionally or physically) while differing in lacking phenomenal consciousness. Sometimes thought experiments about zombies are utilized in philosophy of mind in order to argue, for instance, that consciousness cannot be physically explained, or instead cannot be implemented in a computer.

To give a brief and simplified sketch of a zombie argument, consider the following line of thought that attempts to use the alleged conceivability of zombies in an argument against functionalism (roughly, the view that consciousness can be explained in terms of the sorts of functional causal activities that human brains can have in common with computers).

If consciousness is the sort of thing that can be explained wholly in terms of the functional causal activities in human brains that could be likewise implemented in a robot (activities at a coarse grain that could equally be implemented in a network of neurons and system of microchips) then it ought to be impossible for a creature have exactly that putative set of physical goings-on while lacking phenomenal consciousness. And, let us suppose (a bit contentiously, to be sure), that if something is impossible, then it is inconceivable. However, it seems manifestly conceivable that a robot who acts and talks just like me, and even has a coarse-grained inner causal structure similar to my brain, might nonetheless differ from me in that of the two of us, only I have conscious states.

But given the above suppositions, it would follow then that consciousness cannot be explained wholly in terms of such functional causal activities that might co-occur in both human brains and the computerized control systems of robots. (For a longer discussion of the sorts of issues at play in the present paragraph, see Mandik 2017.)

One of the key pieces in the above line of thought is the supposition that we can coherently conceive of zombies that are behaviourally and functionally identical to normal humans while lacking consciousness. Dennett attacks this supposition by pointing out that it goes hand-in-hand with supposing *zimbo*s to be conceivable. What is a zimbo? A zimbo is a species of zombie invented by Dennett. Recall that a zombie is a creature who is similar to normal humans in some relevant respect, for instance, both a human and his zombie counterpart are imagined to have exactly analogous behaviours, however, zombies are hypothesized to be totally devoid of consciousness. A Dennettian zimbo is a zombie who's key similarity to its human counterpart involves cognition: they have all and only the same cognitive states – thoughts, beliefs, etc. – while differing in that only the normal human has conscious states. Suppose that the human in question is having a conscious visual state of seeing red while also having a conscious feeling of pain. Suppose they have just stubbed their toe on the leg of a red sofa. Suppose further that the human is thinking about these conscious states. The human is thinking a thought expressible as 'I am seeing something red and also experiencing an intense pain.' By definition, the human's zimbo counterpart is thinking an exactly analogous thought about itself. It is thinking a thought expressible as 'I am seeing something red and also experiencing an intense pain.'

Assuming Rosenthalian HOT theory, especially the non-relational reading of it, it follows from the zimbo having that thought that it thereby is in a conscious state. But this is a manifest contradiction of the supposition that the zimbo is a zombie, a creature lacking conscious states. To be sure, the contradiction here is not inherent in the very idea of a zimbo, but instead arises in supposing both that there can be zimbos and that the Rosenthalian HOT theory is the correct analysis of what it means for a creature to be in a conscious state. Either or both must be rejected. Dennett opts to reject both. Dennett uses HOT theory to point out what's absurd about zombies and zimbos, and after rejecting zombies and zimbos, goes on to argue against HOT theory, thus kicking away the ladder he climbed up on.

Dennett's reluctance about the HOT theory seems to have multiple sources. One is a worry that the theory posits more states than there can be evidence for

(1991, p. 319). Another is that the theory seems to invoke a central ‘Cartesian theater’ (1991, p. 320; 2015 pp. 218–19). More about this second point below. Regarding the first point, the HOT theory allows that there can be a real difference, for instance, between the first-order thought of thinking that it is raining and the second-order thought of thinking that one is thinking that it is raining (and so on for third-order, fourth-order, etc. thoughts). Dennett sees no evidence for one’s thinking that one is thinking that it is raining that is distinct from the evidence that one is thinking that it is raining. There are, to be sure, different verbal expressions – there are, for instance, more words spoken in saying ‘I think that I think it’s raining’ than in saying ‘I think it’s raining.’ But these verbal differences, according to Dennett, are not manifestations of pre-existing differences in cognitive state. (For more comparing Rosenthal and Dennett, see Mandik 2015.)

### 3.2 Dennett’s own preferred theory

Just as Rosenthal’s theory of consciousness can be seen as primarily an explanation of what a conscious state is, so can we view Dennett’s theory as primarily an account of state consciousness. Dubbed by Dennett (1991) *the multiple drafts theory of consciousness*, it is the theory that a conscious state is spread out in both space and time in the brain across multiple instances of what Dennett calls ‘content fixations’, each of which – the ‘multiple drafts’ of the theory’s name – compete for domination in the cognitive system. This domination is what Dennett calls ‘fame in the brain’. Dennett sometimes (see Dennett 2005) refers to his theory of consciousness as the *fame in the brain theory of consciousness*. A crucial part of Dennett’s theory of consciousness is the following point: Dennett denies the existence of what he calls ‘the Cartesian theater’ (the central posit common to both the substance dualism of Descartes and a view that Dennett dubs *Cartesian materialism*). Dennett denies both (1) that there’s a single place in the brain where things all come together to give rise to consciousness and (2) that there is a specific time at which the onset of consciousness occurs.

What is the ‘Cartesian theater’ that Dennett is so keen to deny the existence of? The so-called Cartesian theater is where the various previously unconscious brain events march onto the stage of consciousness before the audience of a *homunculus* – Latin for ‘little man’ – who watches the passing show. Dennett regards such a positing of a homunculus as non-explanatory: How is the homunculus conscious of the show in the Cartesian theater?

Dennett's criticism here of the Cartesian theater can be seen as an accusation that his opponents are committing what philosophers call a homuncular fallacy. Consider, for example, attempting to explain perception by positing the creation of mental images that are apprehended with the 'mind's eye'. The problem here is that the alleged explanation threatens an infinite regress. If a person's ability to perceive something is explained by some inner homunculus that itself perceives something, the question arises of how that inner entity is able to perceive anything. Is some additional homunculus, a homunculus within a homunculus, to be posited? Obviously not, for this just infinitely forestalls ever explaining what perceiving is.

Many of the considerations that Dennett provides in support of his multiple-drafts theory of consciousness hinge on the application of a distinction between representational contents and representational vehicles to conscious representations of time. Such representations may themselves (the vehicles) occur at times other than the times that they are representations of (the contents). The importance of the content/vehicle distinction for time representation can be drawn out in contemplation of an argument Dennett gives concerning the illusory motion and illusory colour change in an effect known as the color-phi phenomenon.

In the color-phi phenomenon, the subject is presented with a brief flash of a green circle, followed by a brief flash of a red circle in a different location. Subjects report the appearance of motion. They report seeing a green circle that moves and becomes red. Further, the circle appears to change from green to red at the half-way point of its trajectory.

So, here's a question: How is it that subjects are aware of the green circle turning red before arriving at the spot where the red circle is flashed? The subjects cannot have known ahead of time that a red circle was going to flash, so how is it that they are able to have a conscious experience of something turning to red prior to the red circle's flashing?

In seeking an answer to this question, we may feel pulled toward two candidate explanations, explanations that Dennett has dubbed 'Orwellian' and 'Stalinesque'. Dennett argues against such explanations, and his arguments against them help to further illustrate his cognitive approach to phenomenal consciousness.

One candidate explanation (the Stalinesque explanation, named after Stalinesque show trials) is that the subject unconsciously perceives the red circle's flash and the subject's brain uses that information to generate an illusory conscious experience of a green circle changing to red. The other candidate explanation (the Orwellian explanation, named after the revisionist history promulgated by the totalitarian regime in Orwell's *1984*) is that the subject

consciously perceived only the non-moving green and red circle flashes and has a false memory of there having been a moving and color-shifting circle.

Dennett argues that there is absolutely no basis for preferring one of these candidate explanations over the other. This is plausible, and as I argue in (Mandik 2015, p. 190):

To attempt to persuade yourself of Dennett's conclusion, first imagine being a subject in a color phi experiment. What you introspect is that there has been a visual presentation of a moving, color-changing circle. Your introspective judgment is that you have experienced such an episode. But to resolve the Stalinesque v. Orwellian debate on introspective grounds, your introspective judgment would need to wear on its sleeve whether its immediate causal antecedent was a false memory (Orwellian) or a false experience (Stalinesque). But clearly, no such marker is borne by the introspective judgment. So much for the first-person evidence!

So now, imagine being a scientist studying a subject in a color-phi experiment. Imagine availing yourself of all of the possible third-personal evidence. Suppose you avail yourself to evidence gleaned via futuristic high-resolution (both spatially and temporally) brain scanners. Such evidence, let us suppose, will allow you to determine not only which brain events occur and when, but also which brain events carry which information, and which brain events are false representations. This is, of course, to presume solutions to very vexing issues about information, representation and falsehood, solutions that might beg the question against a Dennettian anti-realism about representation and perhaps, thereby, against Dennettian anti-realism about consciousness, but I won't pursue this line of thought here. However, we will here suppose that such solutions can be arrived at independently of resolving issues about consciousness. Clearly, then, the evidence that you have will, by itself, tell you nothing about which states are conscious. So much for the third-person evidence!

To surmount this hurdle for strictly third-person approaches, you may feel tempted to either ask the subject what their conscious experiences are like, or allow yourself to be a subject in this experiment. However, either way you will only gain access to an introspective judgement with a content that we have already seen as underdetermining the choice between the Orwellian and the Stalinesque.

Given that there's no real difference between the Orwellian and Stalinesque scenarios, what matters for consciousness is what the scenarios have in common, namely the content of the belief or thought that one underwent a conscious experience of a color-changing, moving circle. There's nothing independent of this belief content that serves to make it true, so having a belief with such-and-such content is all there is to being in so-and-so conscious state.

According to Dennett, there is no fact of the matter about consciousness aside from how things seem to the subject, and how things seem to the subject is determined by the belief or thought that is arrived at via the process, smeared out in space and time in the brain, of competitions for fame in the brain via multiple content fixations.

Some of Dennett's critics have accused his argument here of relying on an untenable verificationism, the view that reality does not outstrip our evidence for it. Dennett does not dodge the charge of verificationism, but instead explicitly embraces a sort of verificationism that he calls *first-person operationalism*, a thesis that 'brusquely denies the possibility in principle of consciousness of a stimulus in the absence of the subject's belief in that consciousness' (1991, p. 132). Perhaps our consciousness is just the sort of thing that verificationism or operationalism is totally appropriate for. After all, isn't the main thing that we want explained about consciousness is how our minds seem to us?

Opponents of the cognitive approach may grant that seeming is indeed what is centrally in need explanation in explaining phenomenal consciousness, but may accuse the cognitive approach of conflating two distinct notions of seeming, one of which, a phenomenal sense of seeming, needs to be kept distinct from a cognitive (or 'doxastic') sense of seeming. (For defences of such a distinction, see (Chisholm 1957; Jackson 1977) and (Fred Dretske 1969); for criticisms, see (Gibbons 2005).) Phenomenal seemings viewed as aspects of mental life that are conscious but distinct from one's own cognitive apprehension of them may be viewed as the sorts of things that many have called *qualia*, a sort of mental denizen that Dennett has taken pains to argue against.

### **3.3 Consciousness without Qualia: Dennett's 'Quining' arguments**

The *locus classicus* for Dennett's attack on the very idea of qualia is his famous (1988) article 'Quining qualia'. In this article, Dennett argues against the existence of qualia, where the crucial description of what qualia are supposed to be is that they are properties of consciousness that are (1) intrinsic, (2) ineffable, (3) directly known, and (4) private. Logically, if qualia are correctly defined by that four-part description, and nothing exists that satisfies all parts, then qualia do not exist. In order to establish the non-existence of qualia so-defined, an opponent of qualia need only establish the failure of anything to live up to one part of that description. Dennett, however, goes further, and attempts to cast doubt on all four parts. It seems to be Dennett's view that nothing is intrinsic, nothing is ineffable, and so on. Especially pertinent to the present discussion of

cognitive approaches to phenomenal consciousness are Dennett's attacks on the alleged intrinsicity of qualia and their alleged direct knowability. At the heart of Dennett's case against intrinsicity is his thought experiment of the experienced beer drinker. Against the alleged direct knowability of qualia, Dennett marshals his thought experiment of the coffee tasters Chase and Sanborn.

A property is an extrinsic property if its instantiation depends on the instantiation of other properties. Otherwise, it is intrinsic. Take, for example, the property of being a parent. If no one instantiates the property of being a child, then no one instantiates the property of being a parent. It is a vexing issue of whether any properties are intrinsic. The property of having some particular weight can be shown to be extrinsic, for one's weight would differ if one were on the moon instead of the Earth. It might be thought that mass is intrinsic, since mass can remain the same despite the aforementioned changes in weight. However, in the context of Einstein's Special Theory of Relativity, we can see that mass itself is extrinsic, for the mass of an object increases as its velocity approaches the speed of light, and velocity itself is a matter of motion, which in turn depends on the kinds of relations a body bears to other bodies and perhaps also to space-time itself.

While it is difficult to come up with a clear example of an intrinsic property – a difficulty that may indicate that there actually is no such thing as an intrinsic property – some have thought that so-called qualia are examples of intrinsic properties. One way of conveying the idea that qualia are intrinsic is by reference to the alleged conceivability of intersubjectively undetectable qualia inversions. Conceivably, or so it goes, there could be a being who is behaviourally just like you, including verbal behaviours such as calling out the names of certain colour samples, but what it is like for you to see green is the same as what it is like for them to see red, and vice versa. Suppose further that there is no difference physically between the two of you either. Any inspection of your internal makeup, no matter how fine grained, reveals the same details in the complex arrangements of your physical parts, down to the smallest parts – your cells, molecules etc. No difference detectable from the third-person point of view serves to distinguish you from your doppelgänger, aside from the fact that you can be at different places at the same time. When shown a ripe tomato and asked its colour, you both say 'red'. When asked your favourite colour, you give the same answer. When asked 'which is more similar to red, orange or green?' you both answer 'orange'. Nonetheless, if intersubjectively undetectable qualia inversion is a coherent possibility, then it's possible that what it is like for your doppelgänger to see green, is just like what is like for you to see red. Whereas you

have a red quale in response to a red visual stimulus, your doppelgänger has a green quale in response to the very same stimulus, and a very different stimulus would be needed to elicit a red quale in your doppelgänger. What is supposed to be intrinsic about, for instance, a red quale, is that the red qualia of you and your doppelgänger bear very different relations to your respective behaviours and internal physical structures. Nonetheless, despite all these extrinsic differences, your red quale is just like your doppelgänger's red quale. And thus is a red quale supposed to be intrinsic.

Being an intrinsic property leaves open the possibility that among the relations irrelevant to a quale's nature are any relations it bears to cognitive states. If a red quale is intrinsic, then it has the same internal nature regardless of whether it is a quale in the mind of someone who believes that red is one of the ugliest colours ever or instead believes that it is the most beautiful.

Dennett attacks the alleged intrinsicity of qualia via a thought experiment. Consider a flavour that many consider to be an acquired taste. Many say the flavour of beer is such a flavour, and many adults who love beer recall not having liked it when they first tasted it. If a quale is supposed to be what you apprehend when you apprehend what it is like to have such-and-such experience, the question arises of whether what it is like for you to taste beer is the same now as it was before you grew to like it. Many experienced beer drinkers who underwent a process of coming to develop an appreciation for beer may be tempted to say that if beer tasted like *this* when they first tried it, then they wouldn't have hated it, since *this* flavour (here thought of as a quale), is one that they love. But this line of thinking puts pressure on the idea that what it is like is something intrinsic and in no way dependent on relations, as for instance relations to cognitive states such as states of liking, hating, etc. If what it is like to taste beer while liking it and what it is like to taste beer while hating it are different, then this puts pressure on the idea that what it is like to taste beer is an intrinsic property unrelated to whether one likes it or not.

Dennett turns to another thought experiment for his attack on the alleged direct knowability of qualia. The notion of direct knowing might be conveyed by contrasting it with the indirect way in which many things are known. Consider how it is that you know that you have a brain. It is unlikely that you have ever seen your brain. Even if you had a transparent window in your skull, you would still have to look in a mirror, and draw an inference that the image you see in the mirror correctly reflects the reality of what is in your skull. If you have had an MRI or other kind of medical scan of your brain and seen the resulting images, the knowledge you come to have of your brain is likewise

indirect, for there are inferences that must be made about the reliability of such tests to indicate an underlying reality. For many people, their knowledge that they have a brain is not due to any observation, indirect or otherwise, of their own brain, but instead the result of an inference based on knowledge that they are a human being with typical capacities, and other humans with such capabilities have been shown to have brains. In contrast, so the story goes, your knowledge of whether you currently are experiencing a red quale or a painful quale is supposed to be unmediated by inferences drawn from observations and scientific knowledge. You just introspect and there it is, the quale that is thereby known.

But now let us turn to the sort of question raised by the beer drinker thought experiment: Do qualia have wholly intrinsic natures distinct from any judgements one might have about them? If qualia are directly known, then it is reasonable to suppose that the answer to that sort of question is one that can be known directly. Any dispute should just be settled by directly introspecting one's own mental states. Here is where Dennett's next thought experiment serves to call into question such direct knowability. He invites us to imagine two professional coffee tasters, Chase and Sanborn, who work doing quality control for Maxwell House coffee. Both were hired around the same time, and sought out the job in the first place because they enjoyed Maxwell House coffee. But one day they each confess to the other that they no longer enjoy their jobs because they no longer enjoy drinking that brand of coffee. However, despite these similarities between Chase and Sanborn, they differ in how they characterize their respective mental lives with respect to the predicament they find themselves in. Chase claims that the taste has remained the same, but what has changed is that he doesn't like that taste any more. Sanborn claims that the taste has changed, and claims further that if had remained the same, he would still like it. In referring to the taste, they are not referring to the chemical structure of the coffee. Suppose that that chemical structure has remained the same throughout, and they are both aware of that fact. What they are disagreeing about is whether a mental aspect of their reaction to putting that chemical in their respective mouths has remained the same (as Chase claims) or instead changed (as claims Sanborn). Putting the disagreement in terms of qualia, Chase claims that his coffee-associated quale has remained the same, and what has changed is a cognitive state, a judgement or appraisal that he no longer likes that quale. In contrast, Sanborn claims to now have a different quale associated with drinking coffee: He used to have one that he enjoyed, and would enjoy again if he could regain it, but now he has one that he doesn't like.

Dennett urges that no one, not Chase, not Sanborn and not any third party could settle the dispute between Chase and Sanborn. Suppose we imagine trying to settle the dispute by scanning the brains of Chase and Sanborn, and seeking out some evidence about the brain processing stream wherein information flows from their sensory periphery, through their central nervous system, eventually giving rise to the processing that corresponds to a judgement and ultimately in the musculoskeletal activity that is the expression of that judgement. Imagine that we have been scanning Chase and Sanborn regularly throughout their lives. We might go through the resulting data seeking information about whether what has changed about Chase and Sanborn was something relatively early in the neural processing stream versus something later. There are several problems with seeking to settle the dispute this way. The first is that there's really no way of having any idea which part of the processing stream corresponds to a quale versus a judgement about a quale. But there's another problem, one that is much more directly applicable to the present question of whether qualia are directly knowable. If, in order to settle the dispute about whether what has changed in Chase and Sanborn is not something that Chase and Sanborn themselves know simply by introspecting, but requires instead some third-person accessible data such as brain scans, then the claim that such facts about qualia are directly known is thereby undermined.

#### 4 A challenge to cognitive approaches and a possible solution

One of the main general kinds of complaint against cognitive approaches to phenomenal consciousness is that it requires more conceptual sophistication on the part of conscious creatures than they can plausibly be said to have or employ for every instance in which they are phenomenally conscious. At the core of this complaint is an assumption of a deep connection between cognition and concepts. A typical example of a cognitive state is the thought that there is coffee in the mug in my hand. According to this assumption, in order to think that there is coffee in the mug in my hand, I must both possess and employ concepts such as the concept of coffee, the concept of a mug, relevant concepts of one thing being in another (perhaps different concepts for coffee being in a mug and a mug being in a hand), and so on. A typical example of a state of phenomenal consciousness is my visual experience of a patch of paint as being the most bright and red thing presently in my visual field. Under the assumption linking concepts to cognition, a natural interpretation of the cognitive approach

to phenomenal consciousness is that concepts are going to play crucial roles in my having a phenomenally conscious state. We can cast this point in terms of a HOT-theoretic version of the cognitive approach. If it subjectively seems to me that I am having a visual experience of a paint patch as being bright and red, then I need to have a thought, a higher-order thought, about one of my own visual experiences, and further, I must think of that visual experience as being an experience of something red, not blue, and as something bright, not dark or dull. And having such thoughts, under the assumption being discussed, requires that I have and deploy concepts such as the concept of a visual experience, a concept of red, a concept of brightness and so on.

Here many worries arise hinging on the plausibility of tying phenomenal consciousness to concepts. Perhaps some of the concepts mentioned above, like the concept of a visual experience, are not possessed by, for instance, babies or non-human animals. If such creatures nonetheless have phenomenally conscious visual experiences, then that would contradict at least one version of the cognitive approach, namely the HOT theoretic one just sketched. Another sort of worry in the ballpark, and one that I will dedicate the majority of this section to discussing, is that there is a fine-ness of grain to the contents of phenomenally conscious states that outstrips the concepts possessed and employed by the subjects of such phenomenally conscious states. Focusing specifically on visual experiences of colours, the worry is that even an adult human who possesses concepts of colours – a concept of red, a concept of one colour being brighter than another – there are more colours and aspects of colour than they have concepts for. One particularly famous expression of this ‘fineness of grain’ concern is one due to Evans (1982). Evans’s articulation of the worry has had a wide influence, but he didn’t so much spell out an argument as pose a rhetorical question: ‘Do we really understand the proposal that we have as many colour concepts as there are shades of colour that we can sensibly discriminate?’ (Evans 1982, p. 229). For a full-fledged argument that our experiences of colours have contents that outstrip our conceptual contents, we may turn to the work of Raffman (1995), and examine a powerful empirical argument for the conclusion under consideration. The argument, which Mandik (2012) dubs the Diachronic Indistinguishability Argument (DIA), hinges on (1) a plausible assumption connecting concepts to memory, as well as (2) a widespread empirical phenomenon concerning colours that are discriminable when presented simultaneously but not when presented serially.

To illustrate (2) consider being shown two paint chips that are both shades of blue, though when placed side by side, you can just barely see that one is a slightly

darker shade of blue than the other. If, after the chips had been taken away, you were shown one of them again and asked whether it was the exact same shade of blue as the chip that had been on the right, like most people you would be unable to reliably identify whether this was the same shade of blue. Contrast this situation in which the pair of chips you are initially shown are so different as to differ with respect to hue – suppose that one is a shade of blue while the other is a shade of red. In this situation, like most people you would be quite reliable at correctly re-identifying one of the chips across a short memory delay.

Let us turn now to consider supposition (1) of the DIA, the supposition connecting concepts to memory. Plausibly, the difference in being able to discriminate colours across a memory delay tracks differences in what colour concepts we have. Most English speakers are adept at using basic colour terms like 'blue' and 'red' and it is natural to suppose that they likewise have concepts of blue and red. And further, when the paint chips are so different as one's being red while the other blue, a short memory delay does not disrupt discrimination. For the shades of blue that are extremely similar, plausibly, very few English speakers know the names of those colours, and it may be similarly plausible that they lack concepts for those individual shades of blue.

The gist of the DIA may be put like this: If the cognitive approach is correct, then there shouldn't be more contents to experience than an experiencer has concepts of. When it comes to visible colours, if the colours are one the experiencer has concepts of, as most English speakers have concepts for red and blue, then samples of red and blue are distinguishable across a memory delay. Contrapositively, if a pair of objectively distinct colour samples distinguishable when side by side, call them blue1 and blue2, are not distinguishable across a memory delay, then the subject seems to lack distinct concepts for blue1 and blue2. If we add to the aforementioned assumptions (1) and (2) the additional assumption (3) that the subject has distinct experiences of blue1 and blue2 even when blue1 and blue2 are presented across a memory delay, then there would seem to be more content to visual phenomenal experiences than can be accounted for by the conceptualist resources of the cognitive approach. However, the cognitive approach does have a powerful response to the DIA, and the response hinges on the way a certain conceptualist account can cast doubt on assumption (3) of the DIA.

The average English speaker can describe a wide variety of colours with a relatively meagre vocabulary. One does not need, for example, 40 distinct colour names to describe 40 distinct shades of blue. One may instead employ a combination of comparative colour terms and phrases, for example, 'darker than',

with non-comparative colour terms, for example, 'blue', to describe more shades of blue than one has individual colour names for. An individual ignorant of the colour terms 'navy' and 'cobalt' can nonetheless describe respective samples of them as two shades of blue, the former darker than the latter. Assuming one has concepts corresponding to the terms one is conversant with, we may credit typical English speakers as having comparative and non-comparative colour concepts. In our earlier example of being presented with samples blue1 and blue2 distinguishable when presented simultaneously but not diachronically, adherents of the cognitive approach may describe the relevant experiential content as a conceptualized content expressible as follows: When the samples are presented simultaneously, the colour contents of the experience are expressible as 'two shades of blue, one darker than the other' an articulation of the content that deploys both comparative and non-comparative colour terms. In contrast, when blue1 and blue2 are presented diachronically, in each presentation the subject is only in a position to confidently deploy non-comparative colour terms. They may thus, then, conceive each sample simply as a shade of blue. A crucial component of this cognitive approach to explaining the key data in the DIA is the way that it exploits the indeterminate nature of conceptual contents. In conceiving of a shade simply as a shade of blue, the conceptualization is indeterminate with respect to precisely which shade of blue it is, and can likewise be indeterminate with respect to which other shades it is darker than.

In her presentation of the DIA, Raffman considers whether her opponent can make some sort of appeal to indeterminate representations to neutralize the threat that the argument poses. She objects against all such approaches that while some of our colour concepts are indeterminate, others are determinate and further, there is no introspective differences with respect to degree of determinacy between our experiences of the colours we have determinate concepts of and the experiences of colours we only have indeterminate concepts of. However, it seems that the cognitive approach has a ready reply to Raffman's objection. According to the sort of conceptualism/cognitivism being scouted here, in order for two experiences to seem different to a subject with respect to degree of determinateness the subject must be applying some relevant concepts of degree of determinateness. However, it may very well be the case that typical subjects during typical acts of introspection do not make any such application of a concept of degree of determinateness, either since they lack such a concept or fail to apply it for some other reason, such as not having much practice in distinguishing their own representations with respect to degree of representational determinateness.

There is much more to say than present space permits on the topic of how the cognitive approach can handle the alleged fineness of grain of phenomenal experience, especially as regards colour experience. For further discussion see Mandik (2012, 2013).

## 5 Conclusion

Cognitive approaches to phenomenal consciousness attempt to explain those aspects of mental life in virtue of which, in the Nagelian phrase, there is something it is like, by appealing to a subject's thoughts, judgements, or other cognitive appraisals about their own mental lives. In so doing, such approaches have the promise of offering a reductive or non-circular explanation of phenomenal consciousness by explaining phenomenal consciousness in terms of cognitive states that are not themselves intrinsically phenomenally conscious (by analogy, one might explain water in terms of hydrogen and oxygen, items that individually are not water). The main cognitive approaches I have looked at, the Higher-Order Thought theory of consciousness of David Rosenthal and the multiple drafts or fame in the brain theory of Daniel Dennett, can be seen as opposed to a qualia-centric approach to phenomenal consciousness, where qualia are thought to be aspects of mental lives that are independent of any cognitive appraisals, thoughts or judgements about one's own mental life. Many of the considerations, both pro and con, in the dispute between qualia-centric and cognitive approaches hinge on thought experiments. However, empirical investigations are pertinent, and in particular we have seen the argument due to Raffman (1995) as an attempt to present empirical data about memory and colour discrimination in an attack on approaches to consciousness that would include the cognitive approach. This empirical argument can be conveyed as one concerning whether phenomenal consciousness has a finer grain than can be adequately captured by the conceptually structured states that are definitive of cognition. However, as I have argued, the cognitive approach has the resources to explain the data Raffman appeals to, and thus ward off the threat posed.

## References

- Brown, R. (2015). 'The HOROR Theory of Phenomenal Consciousness' *Philosophical Studies*, 172 (7), 1783–94

- Bruno, M. (2005). A review of Rocco J. Gennaro (ed.) Higher-order theories of consciousness: An anthology. *Psyche*, 11 (6), 1–11.
- Byrne, A. (1997). Some like it HOT: consciousness and higher-order thoughts. *Philosophical Studies*, 86, 103–29.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press.
- Chisholm, R. (1957). *Perceiving: A Philosophical Study*, Ithaca: Cornell University Press.
- Dennett, Daniel (1988), 'Quining Qualia', A. Marcel and E. Bisiach (eds.), *Consciousness in Modern Science*, Oxford: Oxford University Press.
- Dennett, D. (1991). *Consciousness Explained*. Boston, MA: Little Brown, 1–530, Boston, MA: Little, Brown and Company.
- Dennett, D. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, Cambridge, MA: MIT Press.
- Dennett, D. (2015). Not just a fine trip down memory lane: Comments on the essays on Content and Consciousness, in: Munoz-Suarez, C. and De F. Brigard (eds.), *Content and Consciousness Revisited: With Replies by Daniel Dennett*, 199–220, Berlin: Springer.
- Dretske, F. (1969). *Seeing and Knowing*, Chicago: The University of Chicago Press.
- Evans, G. (1982). *The Varieties of Reference*, Oxford: Oxford University Press.
- Frankish, K. (2016) Illusionism as a theory of consciousness, *Journal of Consciousness Studies*, 23 (11–12).
- Frankish, K. (2012). Quining diet qualia, *Consciousness and Cognition*, 21 (2), 667–76.
- Gennaro, R. (2006). Between pure self-referentialism and the (extrinsic) HOT theory of consciousness, in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness*, 221–48, Cambridge, MA: MIT Press.
- Gennaro, R. (2012). *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. Cambridge, MA: MIT Press.
- Gibbons, J. (2005). Qualia: They're not what they seem, *Philosophical Studies*, 126, 397–428.
- Jackson, F. (1977). *Perception: A Representative Theory*, London: Cambridge University Press.
- Jackson, F. (1982). 'Epiphenomenal Qualia', *Philosophical Quarterly*, 32, 127–36.
- Kriegel, U. (2007). Intentional inexistence and phenomenal intentionality, *Philosophical Perspectives*, 21, 307–40.
- Kriegel, U. (2008). The dispensability of (merely) intentional objects, *Philosophical Studies*, 141, 79–95.
- Lycan, W. (1996). *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Mandik, Pete (2016). Meta-Illusionism and Qualia Quietism. *Journal of Consciousness Studies*, 23 (11-12), 140–148.
- Mandik, Pete. (2006). The Neurophilosophy of Consciousness, in Velmans, Max and Schneider, Susan (eds.), *The Blackwell Companion to Consciousness*, 418–30, Oxford: Basil Blackwell.

- Mandik, P. (2009). Beware of the unicorn: Consciousness as being represented and other things that don't exist, *Journal of Consciousness Studies*, 16 (1), 5–36.
- Mandik, Pete. (2012). Color-Consciousness Conceptualism, *Consciousness and Cognition*, 21 (2), 617–31,
- Mandik, Pete. (2013). What is Visual and Phenomenal but Concerns neither Hue nor Shade?, in R. Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience Studies in Brain and Mind*, vol. 6, 219–27, London: Springer.
- Mandik, Pete. (2015). Conscious-state Anti-realism, in Munoz-Suarez, C. and De F. Brigard (eds.), *Content and Consciousness Revisited: With Replies by Daniel Dennett*, 185–97, Berlin: Springer.
- Mandik, Pete. (2017). Robot Pain, in Corns, J. (ed.), *The Routledge Handbook of Philosophy of Pain*, New York: Routledge.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–50.
- Raffman, D. (1995). On the persistence of phenomenology, in T. Metzinger (ed.), *Conscious Experience*. Munich: Imprint Academic Verlag.
- Rosenthal, D. M. (2005). *Consciousness and Mind*, Oxford: Clarendon Press
- Rosenthal, D. (2011). 'Exaggerated Reports: Reply to Block', *Analysis*, 71 (3), 431–7
- Sellars, W. (1997). *Empiricism and the Philosophy of Mind (with an Introduction by Richard Rorty and a Study Guide by Robert Brandom)*, Cambridge, MA: Harvard University Press.
- Velmans, Max and Schneider, Susan, eds. (2006) *The Blackwell Companion to Consciousness*, Oxford: Basil Blackwell.
- Wilberg, J. (2010). Consciousness and false HOTs. *Philosophical Psychology*, 23 (5), 617–38.