The Higher Order Approach to Consciousness is Defunct

Ned Block
NYU

The higher order approach to consciousness attempts to build a theory of consciousness from the insight that a conscious state is one that the subject is conscious of. There is a well-known objection to the higher order approach, a version of which is fatal. Proponents of the higher order approach have realized that the objection is significant. They have dealt with it via what David Rosenthal calls a “retreat” (2005b, p. 179) but that retreat fails to solve the problem.

The fatal objection is short and simple, but some ground-clearing is required in order to state it. There are two quite different approaches to the nature of consciousness. One line of thought (that I have favored) emphasizes the notoriously elusive “what it is like” (Nagel 1974) to have an experience (Block 1978, Chalmers 1996, Levine 1983, McGinn 1989). Higher order theorists often use the phrase “what it is like”, supposing, controversially, that their theories are able to account for what-it-is-like-ness, and I will contest that claim, taking advantage of that common terminology.

The second approach is the one that emphasizes that a conscious state is a state one is conscious of. This perspective, which is the topic of this paper, is supported by one strand of ordinary usage of the term ‘conscious’. As William Lycan (1996, p. 25) says “I cannot myself hear a natural sense of the phrase ‘conscious state’ other than as meaning ‘state one is conscious of being in.’” Perhaps these two approaches are best understood as concerned with two different mental kinds that correspond to two different senses of the term ‘conscious’ (Block 1995, Lycan 2004a, Lycan 2008). Alternatively, perhaps they are best thought of as conflicting approaches to the same thing (Weisberg 2010). A third and I believe better alternative is that they are non-conflicting approaches to the same thing and should be woven together into a single account that emphasizes both perspectives. (See the last paragraph of this article for an attempt in this direction.)

It is however a mistake to suppose that the approach that emphasizes what-it-is-like-ness insists on a characterization of the phenomenon of consciousness as non-relational, i.e. intrinsic in one sense of that term (Dennett 1988, Harman 1990, Weisberg 2010). That characterization derives almost entirely from opponents of the view and should be suspect on that ground alone. Many of the theorists who emphasize what-it-is-like-ness are also scientific realists, and all the major recent accounts of what consciousness is in the brain have been heavily relational (Block 2005b, 2007, 2009).

Higher order approaches divide into those on which the higher order state is perception-like (Armstrong 1968, 1978/1997, Lycan 1987, Lycan 1996, Lycan 2004b) and those on which the higher-order state is a thought (Rosenthal 1986,

---

1997, 2005b, Weisberg 2010). The point I will make applies to both, and perhaps also to higher order “dispositional” theories of consciousness (Carruthers 2000), but except in this sentence, I will be talking entirely about the non-dispositional versions. I will mainly be concerned with the higher order thought version of the view, with occasional mentions of the higher order perception view.

According to (non-dispositional) versions of the higher order approach, what it is for a mental item—a happening or state—to be conscious is for it to be the object of another mental item—*that has a particular causal origin*. This “particular causal origin” is meant to exclude the higher order thought about your sensations that you infer from observing your own behavior or from a brain scan. Sometimes the restriction on origin is spelled out as: not observational or inferential (Lycan forthcoming) or as not involving inference or observation that satisfies a certain access condition (Rosenthal 2005a), a fine point that I will ignore. Thus (collapsing observation into inference for simplicity) the two versions of the higher order approach agree on the following biconditional claim:

**The Higher Order Theory:** A mental state is conscious if and only if the state is the object of a certain kind of representation arrived at non-inferentially

The notion of ‘object of’ is sometimes cashed out in terms of mere accompaniment and sometimes in terms of representation. I shall tend towards the latter formulation. In the case of the higher order thought version, the representation is an assertoric thought.

It should be noticed that when I have a conscious pain, the second order representation in virtue of which the pain is conscious (or perhaps the third order representation in virtue of which the second order representation is conscious) can be expected to be unconscious. Humans do not routinely achieve ever higher and higher orders of representation.

We now come to the most significant preliminary point, one that has not figured in previous formulations of the fatal objection: a higher order theory of consciousness can be held as an immodest (or ambitious) theory that purports to capture what-it-is-like-ness, or as a modest theory of one kind of consciousness, or of consciousness in one sense of the term, higher-order consciousness. The modest version of the higher order approach recognizes another kind of consciousness (or consciousness in another sense of the term), what-it-is-like-ness or phenomenal consciousness.

---

2 “Assertoric” is introduced by Rosenthal in part to try to avoid refutation by the case of unconsciously feeling guilty about, e.g. unconsciously wanting to kill one’s father and marry one’s mother, an unconscious higher order state that is about another unconscious state (Rosenthal 2005b, p. 185). It also is an attempt to avoid a problem raised by Kati Balog (2000a). In one form, Balog’s point is that if the HOT makes for consciousness all by itself, one would not have to bother arranging pleasurable experiences or even imagining them since merely thinking one has them would be just as good. Assertoric thinking is supposedly not so “easy” to do (Rosenthal 2000).
The contrast between the modest and ambitious forms of the view can be illustrated with regard to the following question: why should it be that when you put together an unconscious pain with an unconscious thought about it, you get a conscious pain? On the modest view, this question gets the not very substantive answer that that is what higher order consciousness amounts to, and this answer allows that even when it is not an object of any higher order state, an “unconscious” pain might nonetheless be conscious in the what-it-is-like-ness sense of the term. The ambitious theorist, by contrast, needs a substantive response.

Is the modest version of higher order theory purely terminological? Not quite, since even if one regards it as an account of consciousness in the higher order sense of the word ‘conscious’, there are a number of alternative accounts of that sense of term—for example the higher order thought and higher order perception accounts.

In a recent defense of the ambitious higher order perspective, Josh Weisberg (2010, p. 18) makes much of the claim that the folk “will not consider a state conscious if the subject is in no way aware of it.” But this observation can be explained by the modest thesis that the higher order sense of the term is easy to evoke in folk usage. I believe that it is an unnoticed oscillation between the relatively innocuous modest version and the controversial ambitious version that fuels the fire of the higher order approach.

Rosenthal has been clear in being an ambitious theorist and in the days before Karen Neander’s (1998) version of the objection that I will be elaborating here, Lycan sounded rather like an ambitious theorist. I quoted Lycan (1996, p. 26) on the verbal aspect of higher order approaches, but the point he is making in that passage is really that the verbal point cuts no philosophical ice. He says (p. 26) “Although I cannot myself hear a natural sense of the phrase ‘conscious state’ other than as meaning ‘state one is conscious of being in,’ the philosophical use of “conscious” is by now well and truly up for grabs, and the best one can do is to be as clear as possible in one’s technical specification.”

I believe that the main advantage of the higher order approach is a demystifying perspective on the change from a conscious experience of a pain to an unconscious experience of it (or the reverse change). I have the conscious pain, then I am distracted, and when the pain is not accompanied by any higher order state about it, it is unconscious. However, note that this account is as available to the modest version of the view as to the ambitious version.

The main competitor to the higher order account in its ambitious form (and, arguably in its modest form) has been the “same-order” account, according to which conscious experience is reflexive in that it consists in part in an

---

3 This attitude can be contrasted with Lycan’s “A simple argument for a higher-order representation theory of consciousness” (2001, p. 3-4), in which he gives an argument for a modest form of the view based on ordinary usage. Lycan tells me he was never an ambitious theorist.
awareness of itself (Brentano 1874/1924, Burge 1997, 2006, Byrne 2004, Caston 2002, Uriah Kriegel 2005, Uriah Kriegel & Williford 2006, Levine 2001, 2006, Metzinger 2003, Ross 1961, Smith 1986). For the same-order account to explain the transition between conscious and unconscious states it would have to supply an account of the difference between a pain having and lacking the self-referential property and why this self-referential property comes and goes with consciousness. I believe that this issue is of a piece with the famous “explanatory gap” (Levine 1983), and that, as with the gap, no one has a glimmer of a clue of an idea of how to think about it. (But I don’t take this as a disadvantage of the same-order theory—the higher order account makes consciousness out to be less puzzling than it really is.)

The fatal problem stems from a radical misrepresentation case in which the item that the higher order thought represents, the “target” in Rosenthal’s terminology, does not exist. Can this happen? There are a variety of neurological syndromes in which patients confabulate—what they do or experience is at variance with what they think they are doing or experiencing. A prominent example is anosognosia in which blind patients think they see or paralyzed patients think they are moving the paralyzed arm. In some cases, it appears that the patient is hallucinating, but in other cases the disorder genuinely seems to be a disorder of thought, a matter of “confabulation” (Fotopoulou 2010).

Karen Neander (1998) introduced an example of triplets, each of whom has an assertoric thought expressible as “I have a sensation of green.” One has a sensation of green; another has a sensation of red; and in the empty higher order thought (empty HOT) case, the twin has no relevant sensation at all. Ambitious higher order theorists such as Rosenthal say that what-it-is-like-ness for all three triplets is exactly the same. As Rosenthal (2004, p. 32) puts it,

Suppose my higher-order awareness is of a state with property P, but the target isn't P, but rather Q. …A higher-order awareness of a P state without any P state would be subjectively the same whether or not a Q state occurs. The first-order state can contribute nothing to phenomenology apart from the way we're conscious of it.

Or, as Rosenthal put it more recently (2009b, p. 249):

“If one has a sensation of red and a distinct HOT that one has a sensation of green, the sensation of red may nonetheless be detectable by various priming4 effects. But what it will be like for one is that one has a sensation of green. Similarly if one has that HOT with no relevant sensation at all.”5

4 Unconscious exposure to a stimulus can facilitate later responses to the same or related stimuli. That is unconscious priming.

5 Indeed, Rosenthal thinks that we have evidence of such a determination of what it is like by a HOT independently from any first order state, in a “change blindness” experiment by James Grimes (1996). As Rosenthal (2009a, p. 162) puts it: “In one case of change blindness, a large parrot switches back and forth between being red and green... even when the parrot’s color changes, there is often no change in what it’s like for
What is ambitious about ambitious higher order theories of consciousness is that a state of consciousness is a state of what-it-is-like-ness. However, the ambitious higher order account is committed to the converse as well. We can understand this converse link by considering sensory qualities such as what distinguishes the experience of red from green or pain from thirst. According to some theorists who emphasize what-it-is-like-ness (me for example), these sensory qualities have a kind of consciousness in themselves (namely what-it-is-like-ness--what I call "phenomenal consciousness"). By contrast, higher order theorists think that those sensory qualities can exist without any kind of consciousness. And to preserve the significance of the higher order theory, they say that those unconscious qualities have no what-it-is-like-ness either. I say "preserve the significance" because conscious pain and pleasure matter to us in a way that depends on what-it-is-like-ness; if the putatively unconscious pains had what-it-is-like-ness, it would be a mystery how they could be unconscious. Lacking consciousness requires lacking what-it-is-like-ness and so a state of what-it-is-like-ness is a state of consciousness. As Rosenthal (1997, p. 411) says "...what it's like for one to have a pain, in the relevant sense of that idiom is simply what it's like for one to be conscious of having that pain."

Here is the brief argument I have been building up to. I'll put it in terms of the higher order thought theory, though I believe it applies equally to the higher order perception account. Suppose that at time t, I have an assertoric higher order thought to the effect that I am experiencing seeing something green, but in fact I am having no visual representation at t: the thought is "empty". Let us suppose further that the higher order thought is not arrived at inferentially. Also, I have no other higher order thoughts at t. The theory supplies a necessary and sufficient condition for a conscious episode. An episode is conscious at t if and only if it is the object of an assertoric higher order thought at t, arrived at non-inferentially. The sufficient condition dictates that this thought at t is sufficient for a conscious episode at t. By the necessary condition, that conscious episode at t is the object of a simultaneous higher order thought. In the example, there is only one higher order thought at t, and we can assume it is not self-referential. So there is no conscious episode at t after all. Thus the sufficient condition and the necessary condition are incompatible in a situation in which there is only one non-self-referential higher order representation.

The reason I keep repeating the 'at t' is to emphasize that because there is an appropriate occurrent higher order thought at t, there must be an occurrence of consciousness at t. An occurrence of consciousness must be the object of a simultaneous occurrent thought, but there is only one thought. One

subjects; Grimes’s subjects often continue seeming to see red when the parrot is green.” No doubt someone else has endorsed such a view of “change blindness,” but I haven’t seen it. And for good reason: in standard “change blindness” experiments, items can change color 50 times before the subject notices a change, and it would be a mystery why the subject would settle on one interpretation rather than another to maintain throughout those color changes.
could suppose that it is its own object, but that would make the theory a version of the same-order theory and would not be a defense of the higher order theory. That is the advertised incoherence.6

Weisberg (2010) says what is conscious is always an intentional object, however sometimes the intentional object exists and sometimes not. (If any non-philosophers have gotten this far, the intentional object of a mental state is what it is “about”. Ponce de Leon searched for the Fountain of Youth and so the (non-existent) Fountain of Youth is said to be an intentional object.) One could put Weisberg’s point by saying that there is no such thing as a HOT with no object because there is always an intentional object that is a conscious state, though sometimes a non-existent one! Rosenthal has taken this line though somewhat less explicitly. For example (2002, p. 415) “Conscious states are states we are conscious of ourselves as being in, whether or not we are actually in them.” And (2005a, Ch 7, p. 209): “… HOTs determine what it’s like for one to be in various conscious qualitative states. So erroneous HOTs will in this case result in there being something it’s like for one to be in a state that one is not actually in.”

However—and this is where ambition comes into play—this line of thought does not avoid the argument I gave because it ignores the fact that according to the ambitious form of the view, if a subject has a higher order thought of the right sort at t, then there is something it is like for the subject at t, an episode or occurrence of what-it-is-like-ness at t. No doubt some defenders of the HOT theory will say that in the “empty” cases, what-it-is-like-ness and consciousness are both instantiated only in intentionally inexistent states. If they say this, they reveal that the what-it-is-like-ness they invoke is fake.

To see this, consider the triplets again, but this time let the HOT be the thought that one is in intense, agonizing pain. As before, each of the triplets has a different first order “sensory” state: One triplet has agonizing pain, another has the sensation of tasting chocolate, and the third has no relevant first order state at all. What it is like for all the triplets is exactly the same according to Rosenthal and Weisberg: (“A higher-order awareness of a P state without any P state would be subjectively the same…” (Rosenthal, 2004, p. 132.).) But how can it be that that the triplets all have states that matter in exactly the same way—in this case three states that are bad in themselves (or good in themselves) in the way that real agonizing pain is bad (or that intense pleasure is good)?7 If what-it-is-like-ness is supposed to matter in the same way whether it exists or not, that just shows that ‘what it is like’ is being used in a misleading way. Of course, the higher order theorist could avoid phony ambition by moving to the modest view that the higher order description of the three triplets as having the same higher

6 I say ‘incoherent’ rather than ‘contradictory’ because the theory does not literally contain sentences of the form p and of the form -p, but rather gives conflicting answers about one kind of case.
7 Those who take pain asymbolia (Grahek 2007) as showing that what it is like to have a pain is not bad in itself can substitute the experience of hurting or awfulness in the argument.
order consciousness is compatible with very different kinds of what-it-is-like-ness in the three cases.

Talk of intentional objects and intentional inexistents can serve some legitimate purposes, but such talk can be misleading. This is a case where it misleads.

It will be useful to distinguish between the aboriginal higher order thought view and the new version, produced in response to the empty HOT problem. The aboriginal higher order thought view explained each conscious happening in terms of a pair of actual occurrences, one said to “accompany” or to be “simultaneous” with the other. For example, Rosenthal (1997, p. 741) encapsulates the theory as follows: “The core of the theory, then, is that a mental state is a conscious state when, and only when, it is accompanied by a suitable HOT.” (My computer counted 15 uses of variants of the term ‘accompanied' in this article.) Accompaniment is a straightforward and non-mysterious relation between distinct happenings. Rosenthal (2005b, p. 179) continues to introduce the theory in terms of the attractive idea of an extrinsic relation, as in: “The idea that a mental state’s being conscious is an extrinsic, or relational, property of that state fits strikingly well with the way we actually talk about consciousness.” But he adds a footnote: “There will be reason to retreat on the claim that a state’s being conscious is strictly speaking relational…” And it is the version to which he has retreated, the non-extrinsic relational view that is the source of the trouble.

The aboriginal (extrinsic relational) higher order thought view says (for example) an actual pain is conscious if and only if it is the object of an actual appropriate higher order thought. Someone can hold this view without holding that an empty higher order thought all by itself makes anything conscious. Certainly, the modest higher order theorist can say “The way we talk about consciousness dictates an answer to some questions about which states are conscious, but the presence or absence of consciousness in a case of a higher order thought with no object simply is not decided by the way we talk.” However, this is not a satisfactory response for a higher order theorist who has metaphysical ambitions, especially of the persuasion that I have termed “ambitious”, since it exposes the fact that the modest theorist is not giving a metaphysical account of the nature of consciousness or what-it-is-like-ness. The aboriginal higher order thought view does not have the form of a claim about what consciousness (or what-it-is-like-ness) is, but rather about what the conditions are in which an actual state is a conscious state, leaving it open what to say about the “empty” case. So there is intellectual pressure for the theorist with both kinds of ambition to move from the aboriginal theory to the new one by claiming that a higher order thought with no object can all by itself make for consciousness and what-it-is-like-ness.

I raised the issue earlier of how it can be that an unconscious pain combined with an unconscious thought about it can yield consciousness. The new ambitious view has an answer to this, albeit an implausible one, in terms of some kind of sufficiency: an appropriate higher order thought is somehow sufficient for consciousness (and what-it-is-like-ness) to occur. However, this
ambition is self-destructive, since the new version of the higher order view has to say, as I just noted, that what-it-is-like-ness that is non-existent can matter just as much as what-it-is-like-ness that exists, and that amounts to abusing the notion of what-it-is-like-ness.

There is, however, a way out: to say that what-it-is-like-ness is a property of the HOT itself. That is, the HOT in a case of an empty HOT is the very state that is conscious and has what-it-is-like-ness. Neander’s triplets would now make sense: each is having the same HOT and so no wonder what-it-is-like for them is the same. However, it is not clear how this view can preserve the insight that a conscious state is a state one is conscious of. Perhaps the HOT is always self-reflexive—a state of consciousness of itself. But, as noted earlier, this would be a version of the rival “same-order” theory.

It may be said that the higher order approach can do without what-it-is-like-ness altogether. But the point I have been making applies to consciousness too. If a theory of consciousness is worth anything, it must be about consciousness in a sense that matters to us in the way that conscious agony or ecstasy matters. If a state of being conscious of agony is supposed to matter equally whether it exists or not, the supposed theory of consciousness is worthless.

Thus the higher order thought perspective faces a number of alternatives of which these are the most salient:

1. Adopt a modest view that prescinds from real metaphysical questions about the nature of consciousness. My objection does not apply to this option—but see the paragraph below for an alternative.

2. Swallow the consequence that non-existent conscious agony can be just as bad as existing conscious agony, in which case the view is about consciousness in a merely technical sense of the term.

3. Move to a version of the same-order view.

At the outset I mentioned that the higher order approach assumes that the phenomenon that higher order theories of consciousness are trying to explain is that a conscious state is one that the subject is conscious of. I mentioned an alternative point of view that emphasizes the phenomenal aspect of experience, “what it is like” to have a conscious experience. The two points of view can be usefully combined if we hold that a reflexive conscious state is one that is phenomenally presented in a higher order representation of it (Block 2005a). (This is an ambitious view, but harmlessly so, since phenomenology is simply included in the account.) Since phenomenal presentation of a state requires that it exist, there is no “emptiness” problem. Since this account uses a phenomenal state to represent itself, there is no issue of mistaken representation. Further, the higher order theory (in its ambitious form) is ad hoc (Block 2009, Byrne 1997, Van Gulick 2004). If consciousness is a matter of a relation between a higher order representation and a first order representation, why should it matter what
the causal origin of that higher order representation is? So why is it OK to stipulate that it cannot have arisen through observation (e.g. if I observe my own angry behavior and so come to think about my anger) or inference? If the higher order view in question is a modest theory based in how we use the word 'conscious', it would make sense to read conditions off of usage, but if the theory is meant to—ambitiously-- go deeper into metaphysical reality, the theory should be one for which no ad hoc conditions are required. And this test is passed by the theory that says that a reflexive conscious state is one that is phenomenally presented in a higher order representation of it. The origin of the phenomenal presentation does not matter. A similar point applies to the fact that in the higher order account it has to be stipulated that the higher order representation occurs at the same time (or at least overlaps in time) and in the same mind as the first order representation it represents (Block 2005a, 2009, Van Gulick 2004). No such stipulations are required in the theory that a reflexive conscious state is one that is phenomenally presented in a higher order representation of it, since a state cannot be presented in a representation of it without a kind of temporal overlap, and the presentation relation requires the same mind. The advocate of naturalization and demystification will balk at the use of an unreduced notion of phenomenal, but the naturalistic credentials of the ambitious higher order perspective are themselves suspect given the upshot that conscious agony and ecstasy matter equally whether they exist or not.8

Bibliography


8 I am grateful to Eliza Block, Richard Brown, William Lycan, Michael Tye, Robert Van Gulick and an anonymous reviewer for comments on a previous draft and to David Rosenthal and Josh Weisberg for discussion of the issues. An earlier version of some of the ideas were presented in the NYU Mind and Language Seminar commenting on David Rosenthal in Spring 2010.


