Higher-Order Thoughts, Animal Consciousness, and Misrepresentation: A Reply to Carruthers and Levine

Rocco J. Gennaro

[Chapter Three of Higher-Order Theories of Consciousness, John Benjamins, 2004.]

The higher-order thought (HOT) theory of consciousness has been defended most notably by David Rosenthal (1986, 1990, 1993, 2004, this volume) and also by myself (Gennaro, 1993, 1996). It has also come under attack from several prominent authors over the past few years. In this chapter, I take the opportunity to defend the HOT theory against two of the most important objections raised in the recent literature. The first has to do with animal consciousness, and the second with the charge that HO theories cannot handle cases where the HO state misrepresents the lower-order state.

The standard HOT theory says that what makes a mental state conscious is the presence of an actual (i.e. occurrent) higher-order thought directed at the mental state. For a variety of reasons, I prefer the HOT theory to the higher-order perception (HOP) model, though I have argued that the difference is not as great or as important as has been portrayed in the literature (Gennaro, 1996: 95-101; see also Van Gulick 2000). As will become clear in section 2, I also hold a somewhat modified version of the Rosenthal’s HOT theory.

1. Animal Consciousness and the Dispositionalist HOT Theory (Carruthers)

A. Preliminaries

In his terrific book entitled Phenomenal Consciousness: A Naturalistic Theory, Peter Carruthers (2000) argues for an alternative “dispositionalist” version of the HOT theory of consciousness. I will critique Carruthers’ dispositionalist HOT (DHOT) theory as opposed to the more familiar “actualist” HOT model (AHOT), but my primary focus will be to reply to his treatment of animal consciousness. This section, then, can be thought of as a continuation of a previous exchange on this very topic (Carruthers, 1989, 1992; Gennaro, 1993, 1996, Carruthers 1999, 2000). I should first point out, however, that there is much that I agree with in Carruthers’ book, such as his thoroughgoing “naturalist” approach to consciousness, his criticisms of so-called “mysterians” about consciousness, and his rationale for preferring a HO theory of consciousness to a first-order (FO) theory (e.g. Dretske, 1995; Tye, 1995).

Nonetheless, I continue to disagree strongly with him on animal consciousness.

One of the most common sources of objection to the HOT theory comes from the concern that it rules out or, at least, renders unlikely the possibility of animal (and even infant) consciousness. Indeed, this objection is raised as a matter to routine in the literature (e.g. Dretske, 1995; Seager, this volume), including various introductions to philosophy of mind.
(Kim, 1996: 164-168) It must be taken very seriously, and so I have attempted to respond to it at length in previously published work (Gennaro, 1993, 1996). For example, I have argued that the HOT need not be as sophisticated as it might seem to those who raise this objection. Since most of us believe that many animals have conscious mental states, a HOT theorist must explain how an animal can have the seemingly sophisticated HOTs necessary for conscious states. A simple general argument against the HOT theory might be put in the *modus tollens* form:

(1) If the HOT theory is true, then most animals do not have conscious experiences (because they are incapable of having the relevant HOTs).
(2) Most animals do have conscious experiences.
*Therefore,* (3) The HOT theory is false.

Most HOT theorists, such as myself, are concerned to show that premise 1 is false; that is, we try to show how the HOT theory does not rule out animal consciousness. Peter Carruthers, however, while himself a HOT theorist, rejects premise 2. So what is frequently viewed as an external criticism against the HOT theory is, in this case, an attack from within as far as I am concerned. Carruthers, in short, accepts the HOT theory *and*, without apology, the apparent consequence that animals do not have phenomenal consciousness. Put another way, Carruthers argues as follows:

(1) If the HOT theory is true, then most animals do not have conscious experiences.
(2) The HOT theory is true.
*Therefore,* (3) Most animals do not have conscious experiences.

While most HOT theorists would again reject premise 1, Carruthers believes that this argument is sound. I think it is important, however, to continue to defend the HOT theory against this objection regardless of its origin. We should reject the notion that a lack of animal consciousness follows from the HOT theory. In my view, doing otherwise can only weaken the theory and makes it very difficult to convince others of its truth.

*B. Animal Concepts: A Reply*

It is important first to be clear about Carruthers’ somewhat idiosyncratic terminology. When he speaks of conscious mental states, he uses the term ‘phenomenal’ states and the Nagelian (1974) expression ‘something it is like’ to undergo such mental states. However, Carruthers also speaks of *nonconscious* ‘feels’ due to an ambiguity in the term ‘feel’:

If animal experiences are not phenomenally conscious...then their states will lack *feel*. 
But if the pains of animals, too, lack feel, then doesn’t that mean that animals don’t feel them?...There is no real objection to HOR theory here...merely an ambiguity in the term ‘feel’...The relational property of feeling pain can thus be understood in purely first-order and non-phenomenological terms, just as can the property of seeing red, [but]...we can (and should) deny that the pains which animals feel are phenomenally conscious ones. So we should deny that animal pains have subjective feels to them, or are like anything to undergo...In fact the idea of a feeling of pain which lacks feel is no more problematic than the idea of a percept of red which lacks feel... (2000: 200-201)

We therefore have the key term in the title of Carruthers’ book; namely, ‘phenomenal.’ When he uses that term, he means the kind of conscious subjective feel that is familiar to each of us. He calls it “experiential subjectivity” as opposed to mere “worldly-subjectivity.” Carruthers’ then offers an account of phenomenal consciousness in terms of dispositional HOTs, as we shall see below. He also responds to my (1993, 1996) criticisms of his views on animal consciousness. He first presents the following summary statement of my position as follows: “[i]n order for [mental state] M to count as phenomenally conscious, one does not have to be capable of entertaining a thought about M qua M. It might be enough, [Gennaro] thinks, if one were capable of thinking of M as distinct from some other state N.” (2000: 195) Carruthers then offers the following reply:

What would be required in order for a creature to think, of an experience of green, that it is distinct from a concurrent experience of red?...something must make it the case that the relevant this and that are colour experiences as opposed to just colours. What could this be? There would seem to be just two possibilities. [1] Either the this and that are picked out as experiences by virtue of the subject deploying...a concept of experience, or some narrower equivalent...On the other hand, [2] the subject’s indexical thought about their experience might be grounded in a non-conceptual discrimination of that experience as such... (2000: 195)

Carruthers rejects both possibilities but neither reply is persuasive. He rejects possibility 1 mainly because “this first option just returns us to the view that HOTs (and so phenomenal consciousness) require possession of concepts which it would be implausible to ascribe to most species of animal.” (2000: 195) But Carruthers has once again overestimated the sophistication of such concepts. He mentions concepts such as ‘experience,’ ‘sensation,’ and ‘seeming red.’ But why couldn’t those animal HOTs simply contain concepts more like ‘looking red’ or ‘seeing red’? Is it so implausible to ascribe these concepts to most animals? I think not. Animals
need not have, for example, the concept of ‘the experience of red’ as opposed to just ‘seeing or looking red.’ “I am now seeing red” seems to be a perfectly good HOT.

Similarly, animals need not have HOTs containing the concept ‘experience’ in any sophisticated sense of the term, but why couldn’t they have, say, the concept of ‘feeling’? To use another example, perhaps animals do not have any sophisticated concept of ‘desire,’ but why not some grasp on the related notion of a ‘yearning’ for food. I fail to see why attributing such concepts to the lion chasing the deer would be so “implausible.” Once again, perhaps most animals cannot have HOTs directed at pains qua pains, but why can’t those HOTs contain the related indexical concepts ‘this hurt,’ or ‘this unpleasant feeling’? Having such concepts will then also serve, in the animal’s mind, to distinguish those conscious states from others and to re-identify those same types of mental states on different occasions. In addition, just as there are degrees of sophistication of mental concepts, so there are also degrees of “I-concepts” contained in the HOTs and much the same goes for an animal’s ability to possess them. (Gennaro, 1996: 78-84)

Carruthers then rejects possibility 2 mainly because “this second option would move us, in effect, to a higher-order experience (HOE) account of phenomenal consciousness,” (2000: 196) but he then defers any critical discussion of this alternative until the following chapter eight. I cannot fully address this topic here, but there is room for several brief replies: First, I have already argued (in Gennaro, 1996: 95-101) that the difference between the HOT and HOE [= HOP] models is greatly exaggerated. Others have also questioned this traditional division of HO theories (Van Gulick, 2000), and some have even argued that the HOP model ultimately reduces to a HOT model (Güzeldere, 1995). Thus, Carruthers’ criticism that my view might eventually “move us” to the HOP model is not as damaging as he seems to think. In other words, he has not really replied to my critique; instead, he has at best shifted the debate to the value of the HOT/HOP distinction itself.

Second, after deferring possibility 2 to chapter eight, Carruthers himself also questions the value of the HOP [= HOE] model over and above the HOT model: “The take-home message is: we would never have evolved higher-order experiences (HOEs) unless we already had higher-order thoughts (HOTs); and if we already had HOTs then we did not need HOEs. Upshot: if we are to defend any form of higher-order representation (HOR) theory, then it should be some sort of HOT theory...” (2000: 219) Whether or not Carruthers is right about this, it clearly does not address possibility 2 left over from his previous chapter. Moreover, he seems now to be dismissing the value of the HOP view in favor of the HOT model, without returning to his objection to my view. On the other hand, as Carruthers (this volume) makes even clearer, he understands his HOT theory to be a form of HOP theory, and so it is again difficult to see why any “move” in that direction would be so problematic for me.
Third, when Carruthers speaks of “non-conceptual discrimination” in, say, one’s perceptual field, it seems to me that this is very misleading or, at least, rather ambiguous. It is thus open to a similar counter-reply to possibility 1. Such experiences may not include the ability to apply some concepts, but surely at least some other less sophisticated concepts are still required in order to have the experience itself. An infant or dog may not experience the VCR as a VCR, but they surely at least apply some concepts to that visual experience, e.g. black and rectangular. Similarly, an animal or infant may not be aware of its desire to eat qua concept of desire, but they can still be aware of that mental state in virtue of some other related concepts, e.g. ‘yearning to eat something.’ I believe that conscious experience always involves the application of some concepts, though I do recognize that this has become a hotly contested claim. I believe that the idea of non-conceptual experience is, at the least, a somewhat misleading and ambiguous notion. (For more on the literature on nonconceptual content, see Gunther, 2003.)

Fourth, in my 1993 and 1996 replies to Carruthers, I was careful not to rely solely on the conceptual considerations he cites. I also put forth behavioral, evolutionary, and comparative brain structure evidence for the conclusion that most animals are conscious. For example, I explained how many lower animals even have some kind of cortex, not to mention the fact that they share with us many other “lower” brain structures known to be associated with conscious states in us. While Carruthers is very knowledgeable about brain science and discusses evolution elsewhere in his book, his failure to do so in the context of this disagreement is significant because the cumulative effect of such strong inductive evidence in favor of animal consciousness is lost. It is very puzzling why a thoroughgoing naturalist would not take such collective evidence to outweigh any considerations which may or may not follow from the HOT theory.

Finally, as Lurz (2002: 12-13) explains, it seems unlikely that the belief that many animals have conscious states can be so easily explained away, as Carruthers tries to do, merely as some anthropomorphic process of “imaginatively projecting” what it is for us to have certain mental states onto various animals. For one thing, it certainly does not describe what goes on in me when I attribute mental states to animals. I agree with Lurz that my main initial reason for believing that animals have conscious mental states has more to do with the fact that their behavior is best explained and predicted by attributing such folk psychological notions to them. In addition, as I mentioned above, such a conclusion is only further supported upon examination of the scientific brain evidence.

C. Brain Structure and Evolution

With regard to brain structure, Carruthers is mainly concerned to discredit the AHOT theory with what he calls the “objection from cognitive overload.” (2000: 221). Indeed, this is
Carruthers’ main objection to the AHOT theory:

...a major problem with the actualist version of HOT theory...[is] the implausibility of supposing that so much of our cognition should be occupied with formulating and processing the vast array of higher-order thoughts necessary to render our experience conscious at each moment of our waking lives... (2000: 221)

But, first, it is never made clear why AHOTs would take up so much “cognitive space” (i.e. neural mechanisms). No neurophysiological evidence is offered to show that our brains aren’t “big enough” to handle the job. After all, it is not just the number of neurons in our brains, but also the numerous connections between them. Second, it is unclear just how Carruthers’ own DHOT theory would avoid this problem (assuming there is one), since even he acknowledges the presence of various actual brain structures (e.g. theory of mind mechanism) which are required to fill out the DHOT theory. For example, Carruthers is forced to acknowledge that his theory requires the presence of “something categorical [= actual] taking place in me whenever I have a conscious experience, on the [DHOT model] -- the perceptual contents are actually there in the short-term memory store C, which is defined by its relation to HOT.” (2000: 233) Rowlands (2001: 305-309) makes a similar point and then goes on to argue, convincingly in my view, that if these actual structures play such an important role in making mental states conscious, then the result is, in effect, to abandon the DHOT view. Dispositional states, on at least some interpretations, also require similar actual brain structure or “cognitive space.” This is not of course to say, however, that I agree with Rowlands’ unrelated criticisms of the AHOT theory.

Of course, part of the reason that Carruthers is so convinced by this argument against the AHOT theory has to do with his acceptance of what he calls the “richness of phenomenally conscious experience” (2000: 299ff; cf. 224) which is often related to a belief in nonconceptual content of experience mentioned earlier. This is a major topic that I cannot pursue here, except to note that I think a very strong case can be made for the conclusion that the phenomenally conscious aspects of our experiences are frequently much less “rich” than is commonly supposed (see e.g. Dennett, 1991; Weisberg, 1999). As Carruthers himself discusses, objects in the periphery of our visual field seem to lack the kind of rich determinacy that Carruthers seems to have in mind. But even within areas of one’s visual focus (e.g. looking at a large painting), a case can be made that one’s conscious attention is more fragmented than is often believed. Therefore, the HOTs in question again need not be as complex as Carruthers seems to think.

With regard to evolution, Carruthers tells us that he finds no evolutionary reason to suppose that actual HOTs be present in the case of conscious mental states: “What would have been the evolutionary pressure leading us to generate, routinely, a vast array of [actual] HOTs
concerning the contents of our conscious experience?” (2000: 225). But, I suggest, that there are at least three good reasons overlooked by Carruthers: (1) On the AHOT theory, actual nonconscious HOTs, we may suppose, can more quickly become conscious HOTs resulting in introspective conscious mental states. On the AHOT theory, introspection occurs when a nonconscious HOT becomes conscious and is thus directed internally at another mental state. The ability for an organism to shift quickly between outer-directed and inner-directed conscious states is, I believe, a crucial practical and adaptive factor in the evolution of species. For example, an animal that is able to shift back and forth between perceiving other animals (say, for potential food or danger) and introspecting its own mental states (say, a desire to eat or a fear of one’s life) would be capable of a kind of practical intelligence that would be lacking otherwise. It seems reasonable to suppose that such quick transitions are more easily accomplished by having actual HOTs changing from being unconscious to conscious. (2) Even if we suppose that some lower animals are only capable of first-order conscious states (and thus only nonconscious HOTs), the evolutionary foundation has been laid for the yet more sophisticated introspective capacities enjoyed by those of us at the higher end of the evolutionary chain. Thus, the presence of actual (nonconscious) HOTs can be understood, from an evolutionary perspective, as a key stepping stone to the capacity for introspective consciousness. Such an evolutionary history is presumably mirrored in the layered development of the cortex. (3) Finally, as Rolls points out, having actual HOTs allows for the correction of plans that result from first-order processing. Rolls puts forth his own modified version of an AHOT theory, and suggests that “part of the evolutionary significance of this type of higher-order thought is that it enables correction of errors made in first-order linguistic or in non-linguistic processing.” (Rolls, 1999: 249)

D. The Moral Issue

Linking his discussion of animal consciousness back to moral issues, Carruthers explains that he had previously argued “that non-conscious pains – pains which would lack any subjective qualities, or feel – could not be appropriate objects of sympathy and moral concern.” (2000: 203; cf. Carruthers 1989, 1992). But Carruthers has had a change of heart. He first imagines a conscious, language-using, agent called Phenumb “who is unusual only in that satisfactions and frustrations of his conscious desires take place without the normal sorts of distinctive phenomenology.” (2000: 206; see also Carruthers, 1999). Without becoming bogged down in the details of Carruthers’ questionable thought experiment, he ultimately argues that Phenumb is an appropriate object of moral concern and that the example shows “that the psychological harmfulness of desire- frustration has nothing (or not much) to do with phenomenology, and everything (or almost everything) to do with thwarted agency.” (2000: 207) In essence, Carruthers is attempting to separate desire frustration from consciousness in order to make room
for the idea that animals can indeed be the objects of sympathy and moral concern, contrary to his previously held position. He explains that his “present view is that it is first-order (not necessarily-phenomenal) disappointments and frustrations of desire which are the most basic objects of sympathy and (possible) moral concern. And these can certainly be undergone by many species of non-human animal.” (2000: 205)

I am frankly very puzzled by Carruthers’ view here for several reasons. First, it seems to me that any actual organism (that we know of) capable of “desire frustrations” will also be capable of phenomenally conscious pains and would thereby also have the ability to suffer. Even the hypothetical Phenumb begins as a conscious agent. It seems to me that desire frustration is an even more sophisticated and intellectual psychological capacity than the mere ability to subjectively feel pains. Even if the two capacities are somehow theoretically distinct, I fail to see what positive reason we could ever have to attribute only the former to any known animal. Second, when Carruthers speaks of “desire frustrations,” it is unclear to me how could they be non-conscious at all. I am not sure that I even understand the idea of a non-phenomenal “disappointment” or “desire frustration.” As Cavalieri and Miller (1999: 3) put it, “[s]o nonhumans can be disappointed, can have their desires frustrated...but these disappointments and frustrations are non-phenomenal. This is just incoherent...Disappointments and frustrations are possible only for the sentient.” Of course, there can be non-conscious desires (and even, pains), but it does not follow that there are non-conscious desire frustrations, especially in organisms who are supposed to be utterly (phenomenally) non-conscious.1 Third, even if we can imagine the possibility of some Spock-like character only able to have such purely intellectual frustrations (as Carruthers suggests in his 1999: 478 fn. 26), it does not follow that such frustrations would be entirely non-phenomenal. They may be devoid of the typical accompanying emotions, but there must, at minimum, be conscious thoughts and beliefs about the objects of those desires. Carruthers is now curiously and suddenly comparing (what he takes to be) non-conscious animals to a very sophisticated intellectual hypothetical character. Thus, in the end, I do not believe that Carruthers’ current moral stance is any more tenable than his previously held view.2

E. Against the Dispositional HOT Theory

Although my main concern has been to reply to Carruthers on the topic of animal consciousness, I will conclude this section with a few critical observations of his DHOT theory as such. (1) It seems to me that Carruthers is not always careful in the way that he phrases his theory. Perhaps best is when he says that “[i]n contrast with the actualist form of HOT theory, the HOTs which render M conscious are not necessarily actual, but potential...There need not actually be any HOT occurring...” (2000: 227) There is, of course, much more to Carruthers’ theory and I cannot hope to do justice to all of the subtleties here (see especially his chapters
eight and nine), but we should at least note the following elaboration of his view:

We can propose that conscious experience occurs when perceptual contents are fed into a special short-term buffer memory store, whose function is *inter alia* to make those contents available to cause HOTs about themselves, where the causation in question will be direct, not mediated by any form of inference. (2000: 228)³

According to Carruthers, then, it would seem that the HOTs in question are not actual HOTs directed at a mental state M, but dispositional HOTs. However, there is significant ambiguity in the way that Carruthers phrases his view. Consider, for example, his slogan at the end of the book: “A disposition to get higher makes consciousness phenomenal.” (2000: 329) Of course, like most slogans, it may necessarily be somewhat oversimplified and even misleading, but it seems to me that the problem runs deeper. The slogan might suggest that the disposition in question belongs to the *mental state M* instead of the HOT. What exactly has the “disposition to get higher”? If the DHOT is some form of HOT, then the HOT is, after all, already “higher” in some sense. This would seem to leave the *lower-order* state as having the disposition in question, which does not seem to be Carruthers’ considered view nor does it seem to make very much sense. One might be inclined to dismiss this problem in the context of a mere slogan; however, in one of Carruthers’ more formal statements of DHOT theory, he says the following:

Any occurrent mental state M, of mine, is conscious = M is disposed to cause an activated belief (possibly a non-conscious one) that I have M, and to cause it non-inferentially.” (2000: 227)

I find this definition unclear and, again, perhaps even at odds with what is supposed to be the DHOT theory. Carruthers is here clearly attributing the disposition in question to M, not to any higher-order state. M becomes conscious when it has a disposition to cause some kind of higher-order state directed at M. This not only contradicts other statements of the DHOT theory, but it is also very difficult to understand on its own terms. At minimum, some clarification is needed.

(2) There is also a crucial distinction in the AHOT theory which seems lost, or at least unaccounted for, on Carruthers’ DHOT theory. Indeed, in Gennaro 1993, I criticized Carruthers for conflating this distinction which, in turn, led him to some of his conclusions regarding animal consciousness. This well-known distinction, mentioned briefly earlier, is the difference between first-order (i.e. world-directed) conscious mental states and introspective (i.e. inner-directed)
conscious states. On the AHOT theory, the former will be accompanied by nonconscious HOTs; in the latter case, there will be conscious HOTs accompanied by yet higher (third-order) nonconscious HOTs. Now this distinction is noticeably absent from Carruthers’ alternative DHOT account, except perhaps for one very brief mention of third-order states (2000: 251-252). Very little is said about this critical difference in conscious mental states. It is therefore left unaccounted for and unclear on his DHOT theory. My sense is that Carruthers is once again conflating, or just ignoring, this important distinction. For example, Carruthers speaks of focusing “on my experience of a colour...and this is to focus on the subjectivity of my experiential state.” (2000: 184) This suggests introspective consciousness, not just a first-order conscious color experience accompanied by an nonconscious HOT. One reason that this is so important is that this conflation also leads erroneously to the conclusion that animals cannot have conscious mental states. If, for example, one mistakenly supposes that the HOT theory requires introspective states to accompany first-order conscious states, then one may very well doubt the possibility of animal consciousness. In any case, Carruthers needs to answer the following questions: How does he explain the difference between first-order conscious and introspective conscious states on the DHOT model? Are the HOTs potential (or dispositional) HOTs only in the former case? If so, do they become actual conscious HOTs in the introspective case? If not, then how can the DHOT model account for the difference? Would there be an additional level of dispositional HOT in the introspective case? Whether Carruthers can answer these questions in a satisfying way without making major modifications to, or even abandoning, his DHOT theory is, in my opinion, very doubtful. Perhaps he can.

(3) Part of the reason for Carruthers’ attack on animal (and infant) consciousness has to do with his allegiance to the so-called “theory of mind” theory, whereby understanding mentalistic notions presupposes having a “folk-psychological” theory of mind (see Carruthers and Smith, 1996). Once again, however, Carruthers builds a great deal into having such a theory and then explicitly ties it to the capacity for having any HOTs at all. For example, he explains that “...the evidence is that children under, say, the age of three lack the concepts of appearance or seeming – or equivalently, they lack the idea of perception as involving subjective states of the perceiver – which are necessary for the child to entertain higher-order thoughts about its experiences.” (2000: 202) But, once again, Carruthers seems to have in mind conscious HOTs, which are (once again!) not necessary for having first-order conscious states according to the AHOT theory. He also soon thereafter makes a similar point about autistic people who have been thought to be “mind-blind” in certain ways (Baron-Cohen, 1995). But then Carruthers goes on to make the same error by telling us that “if autistic subjects are blind to their own mental states, then that will mean they are incapable of self-directed HORs; which in turn will mean that they lack phenomenally conscious mental states, if any form of HOR theory
is correct.” (2000: 202, emphasis added) If by “self-directed HORs” Carruthers means “introspective states,” then he is plainly mistaken about what follows.

In line with many ‘theory-theorists,’ Carruthers also holds that animals with HOTs should be able to have HOTs about the mental states of other creatures as, for example, we might expect to find when animals engage in deceptive behavior. But even if some or most animals cannot engage in deceptive behavior and so do not have HOTs about the mental states of others, it still does not seem to follow that they cannot have HOTs about their own mental states (Ridge, 2001). I therefore agree that Carruthers’ view rests on the false assumption “that there could not be an agent capable of having HOTs about its own mental states but incapable of having HOTs about the mental states of others.” (Ridge, 2001: 333) One might still believe, with Kant, that having HOTs presupposes some sort of implicit “I-thought” which distinguishes the thinker from outer objects (Gennaro, 1996: 78-84 & ch. 9), but those outer objects need not always include the mental states of other conscious beings.

Overall, then, I believe that the AHOT theory remains the leading HOT theory, and it is consistent with the view that most animals have conscious mental states.

2. Misrepresentation and the Division of Phenomenal Labor  (Levine/Neander)

Joseph Levine (2001) raises an important objection to all higher-order theories of consciousness. He credits Karen Neander (1998) for an earlier version of this type of objection under the heading of the “division of phenomenal labor.” The main idea behind the objection is that when “we are dealing with a representational relation between two states, the possibility of misrepresentation looms.” (Levine, 2001: 108; cf. Neander, 1998: 418ff.) Levine then argues that the HOT theory cannot explain what would occur when the higher-order state misrepresents the lower-order state. The main example used is based on color perception, though the objection could presumably be extended to other kinds of conscious states. Levine says:

Suppose I am looking at my red diskette case, and therefore my visual system is in state R. According to HO, this is not sufficient for my having a conscious experience of red. It’s also necessary that I occupy a higher-order state, say HR, which represents my being in state R, and thus constitutes my being aware of having the reddish visual experience...Suppose because of some neural misfiring (or whatever), I go into higher-order state HG, rather than HR. HG is the state whose representation content is that I’m having a greenish experience, what I normally have when in state G. The question is, what is the nature of my conscious experience in this case? My visual system is in state R, the normal response to red, but my higher-order state is HG, the normal response to being in state G, itself the normal response to green. Is my
Levine (2001: 108) initially rightly points out that we should reject two possible answers, each of which is very problematic (cf. Neander, 1998: 420ff):

Option one: The resulting conscious experience is of a greenish sort.
Option two: The resulting conscious experience is of a reddish sort.

Options one and two are both arbitrary and poorly motivated. Even worse, option one would make it seem as if “the first-order state plays no genuine role in determining the qualitative character of experience...” (Levine, 2001: 108) The problem here is that one wonders what the point of having both a lower and higher-order state would be if only one of them determines the conscious experience. On the other hand, if we choose option two, then we have the same problem, except now it becomes unclear what role the higher-order state plays. It would then seem that higher-order states are generally not needed for conscious experience, which would also be disastrous for any HO theorist. Thus, Levine says later on: “When the higher-order state misrepresents the lower-order state, which content – higher-order or lower-order – determines the actual quality of experience? What this seems to show is that one can’t divorce the quality from the awareness of the quality.” (Levine, 2001: 168, italics added)

However, both Levine (2001: 108-109) and Neander (1998: 429-430) recognize that other options are open to the HO theorist, but then dismiss them as well. I will focus on Levine’s treatment of these alternatives, and argue that these options are more viable than he thinks and that they are also closely related.

Option three: “...when this sort of case occurs, there is no consciousness at all.” (Levine, 2001:108)

Option four: “A better option is to ensure correct representation by pinning the content of the higher-order state directly to the first-order state.” (Levine, 2001: 108)

First of all, it is a bit unclear what Levine means, in option three, by “no consciousness at all.” Presumably, he does not mean that the hypothetical person in question would be completely unconscious. This would be a very unnecessary and puzzling consequence of any HO theory. It would also be to confuse “creature” consciousness with “state” consciousness, to use Rosenthal’s terms. So it would seem that Levine’s option three is really saying that, in such cases of misrepresentation, the person has neither the greenish nor the reddish conscious experience. But then it becomes unclear why Levine rejects option three as ad hoc (2001: 108).
What exactly is so *ad hoc* about that reply? The HOT theory says that when one has a conscious mental state \( M \), it is accompanied by a HOT that “I am in \( M \).” If there isn’t a “match” between the contents of the lower-order and higher-order states, then it seems perfectly appropriate for the HOT theorist to hold that something like option three is a legitimate possibility. After all, this is an abnormal case where applying the HOT theory could not be expected to result in a normal conscious state. We are not told just how unlikely or abnormal such a scenario would be. There is an important lack of detail in Levine’s thought experiment; recall that we are simply told to “suppose because of some neural misfiring (or whatever)...”. Perhaps there would be no resulting conscious experience of the diskette case at all.

Alternatively, if there are certain brain lesions involved, perhaps there would be some kind of loss of color vision (achromatopsia) with respect to the diskette case. The diskette case might then be experienced as neither green nor red.

This brings us to the “better option” in option four. In a sense, then, a defense of option three leads us naturally into option four. Indeed, they seem to be two sides of the same coin because defending option three is, in essence, arguing that a match between the higher-order and lower-order states must be “ensured” in order to result in a conscious experience with respect to the relevant concepts involved, e.g. “by endowing [the content of the higher-order state] with demonstrative content.” (Levine, 2001: 108)

Levine does mention two problems with this fourth approach, but I am very puzzled by his remarks. He first asks “what if the higher-order state is triggered randomly, so that there’s no first-order sensory state it’s pointing at? Would that entail a sort of free-floating conscious state without a determinate character?” (Levine, 2001: 109) The answer to the second question is clearly no because, in that case, you would have merely an unconscious HOT without a target state. An unconscious HOT, by itself, cannot result in a conscious state of any kind. Second, Levine simply expresses puzzlement about just how option four “overcomes the basic problem...[which] is that it just doesn’t work to divide phenomenal labor.” (Levine, 2001: 109) This is not really another objection to option four or to the HOT theory; it merely repeats Levine’s conclusion.

In addition, to revisit the initial thought experiment, when Levine says that my “visual system is in state \( R \)...but my higher-order state is \( HG \),” this is highly misleading and perhaps even begs the question against the HOT theory. What encompasses the “visual system”? Levine assumes that it is only the lower-order state \( R \). However, if the HOT theory is true, it seems much more plausible to treat the *entire* system (including both the lower-order and higher-order state) as parts of the “visual system” in this case. Thus, the visual system (or at least the conscious visual system) would have to contain \( R \) and \( HR \), so that there would be a conscious reddish experience (even if an idle \( HG \) state also exists). Perhaps option two is thus not so arbitrary after all. If so, then the hypothetical scenario would seem to be misdescribed or
just assumes the falsity of the HOT theory. HOTs should be understood as part of the “visual system” when one is having a conscious perception of any kind. We should also say, then, in this case, that R and HR are each necessary for having a reddish experience, but neither one is sufficient by itself. R and HR are jointly sufficient. As I will urge later, empirical evidence also exists for such a characterization of the HOT theory.

It might be useful here to contrast Levine’s example with two other abnormal cases. (1) In cases of visual agnosia subjects suffer from the inability to recognize perceived objects, but the disorder is neither due to any intellectual deficiency nor to any basic sensory dysfunction. The visual agnosic consciously perceives things, but cannot recognize what they are. They often mistake one object for another; for example, even mistaking one’s wife for a hat (Sacks, 1987). This differs from Levine’s case in that the mere perception of objects remains intact, including the color perception. Visual agnosics are not blind and do not have damage to area 17 of the visual cortex. However, a HOT theorist might view this case as one where a HOT does not “match up” with the first-order visual input. Thus, it seems reasonable to view this as a case where the “normal” concept in the HOT does not accompany the input received through the visual modality (Gennaro, 1996: 136-138). (2) On the other hand, there is also the well-known phenomenon of blindsight (Weiskrantz, 1986, 2000) whereby patients can sometimes accurately answer questions about objects without having any conscious perceptions of those objects, such as detecting the movements and shapes of objects in the blind visual field. Unlike Levine’s case and the visual agnosic, the blindsight patient does not have a conscious perception at all of objects in her visual field (due to lesions of the visual cortex), though the common wisdom is that some visual information is processed in other “secondary” parts of the brain in ways that explain the behavioral data. In this case, there would be no HOT at all directed at the lower-order visual input because there is no conscious perception at all. At best, we have a case of a nonconscious mental state or an informational state influencing behavior (Gennaro, 1996: 129-134).

Nonetheless, I do think that Levine and Neander have, in a somewhat indirect way, hit upon one very important and potentially troubling issue regarding the nature of HOT theory. There may indeed be an element of truth in Levine’s argument; namely, that it is difficult to make sense of entirely splitting off the lower-order state from the HOT. Thus, I do believe that he is grappling with a deeper issue that must be addressed by any HOT theorist. It is perhaps best expressed by Levine when he says that the HOT theory has a difficulty with

the paradoxical duality of qualitative experiences: there is an awareness relation, which ought to entail that there are two states serving as the relevant relata, yet experience doesn’t seem to admit of this sort of bifurcation. Let’s call this the problem of “duality.”
The problem of duality is an important problem, but I think it can be handled best by adopting a variation of Rosenthal’s HOT theory. According to the variation that I have defended elsewhere (Gennaro, 1996), first-order conscious mental states are complex (or global) states comprised of both the lower and higher-order states. Consequently, I think that consciousness is an intrinsic and essential feature of conscious mental states, unlike Rosenthal who holds that the HOT is an inessential and extrinsic property of conscious states. I have called this position the “wide intrinsicality view” (WIV) and have argued for its superiority over Rosenthal’s model (Gennaro, 1996: 24-30). In this context, though, I believe that the WIV can help to alleviate some of the puzzlement expressed by both Neander and Levine. For example, in the quote above, a proponent of the WIV can respond that conscious experience (from the first-person point of view) does not seem to allow for a split (“bifurcation”) between the lower-order and higher-order states. However, the “awareness relation” still does not entail the existence of two entirely separate states. Instead, on the WIV, we have two parts of a single conscious state with one part directed at (“aware of”) the other. In short, there is a complex conscious mental state with an inner intrinsic relation between parts. There is, therefore, a kind of “self-referential” element in consciousness. This general idea is most closely linked to Brentano (1874/1973), but I have also argued at length that it is the best way to understand Sartre’s theory of consciousness (Gennaro 2002).

Moreover, this variation allows us to avoid one controversial aspect of Rosenthal’s theory which is also a target of Levine’s critique: Rosenthal’s theory “splits off subjectivity from qualitative character, and ... it is precisely this feature that seems so implausible.” (Levine, 2001: 105). Unlike Rosenthal (1991), I do not hold that the there can, for example, be unconscious sensory or qualitative states. Some of the disagreement here is purely terminological, e.g. how to use the terms ‘sensory,’ ‘experience,’ and the like. However, there is also substantial disagreement. Since Rosenthal believes that HOTs are extrinsic and inessential to their target states, he (unlike me) holds that the lower-order states can exist without HOTs and continue to have their qualitative properties. On my view, however, it is the HOTs which bring the intrinsic qualitative properties into the conscious state. So, for example, on my view there can be unconscious pains and perceptions, but they are not ‘sensory’ or ‘qualitative’ states while unconscious. When a pain or perception becomes conscious by virtue of becoming the target of an appropriate HOT, it then becomes a qualitative state.

In a striking passage, Neander credits Barry Loewer for an “ingenious suggestion that might be worth pursuing” (1998: 430):
the suggestion is that the two levels of representation might be collapsed into one level that is self-referential....This suggestions also rids us of the division of phenomenal labor, while still allowing us to maintain that the difference between conscious and unconscious sensory representations is that some of them are meta-represented and some are not. Since the first and second-order representings no longer involve two separate representations...the two cannot come apart, so mis-(meta-)representation is in principle impossible. (Neander, 1998: 429-430)

This sounds very familiar, but Neander unfortunately also dismisses this option too quickly (1998: 430). I hope I have shown that this is a truly viable option which can help to counter what Levine calls “the problem of duality” along the lines of his option four.

Furthermore, it is also now possible to address some of Levine’s other concerns about whether or not qualia can be explained as either intrinsic or relational features of conscious states (2001: 93-107; cf. Levine 1995). With the WIV alternative, we can now see that this is a false dichotomy. Qualitative states can be complex states with both intrinsic and (inner) relational features. This solution is perhaps similar to what Levine calls “the complexity gambit,” (2001: 95) but I have already argued that such a move can address his challenge about the nature of qualitative states while, at the same time, not admitting “that no progress can be made [in explaining consciousness] if we consider qualitative character to be an intrinsic property of experience...” (Levine, 2001: 94) This is yet another advantage of the WIV over Rosenthal’s HOT theory: Consciousness can be both an intrinsic and relational property of experience without giving up on any further explanation of its nature. A case could be made that Rosenthal mistakenly infers that treating consciousness as an intrinsic quality of experience forces one into an allegiance with the unsatisfying Cartesian position whereby consciousness is an unanalyzable property of conscious states (see Gennaro, 1996: 21-24).

Finally, I wish briefly to mention some very suggestive and relevant empirical evidence which I think supports the HOT theory in general and the WIV in particular. Gerald Edelman and others have argued that loops (or re-entrant pathways or back projections) in the neural circuitry of the brain are essential for conscious awareness (e.g. Edelman and Tononi, 2000). As Churchland puts it, “the idea is that some neurons carry signals from more peripheral to more central regions, such as from V1 to V2, while others convey more highly processed signals in the reverse direction...it is a general rule of cortical organization that forward-projecting neurons are matched by an equal or greater number of back-projecting neurons.” (Churchland, 2002: 148-149) I cannot go into further neurological detail here, but it is worth making the following observations in this context: (1) The brain structures involved in loops seem to resemble the structure of at least some form of HOT theory; namely, that lower-order and higher-order states
are combining to produce conscious states. (2) The importance of higher-order *concepts* in conscious experiences is also readily apparent. Part of the reason why Edelman and others believe that back projections play a prominent role in consciousness is that “perception *always* involves classification; conscious seeing is *seeing as.*” (Churchland, 2002: 149) This is a key aspect of any HOT theory. (3) More specifically, such evidence seems to support the WIV version of the HOT theory because of the intimate and essential relationship between the “higher” and “lower” areas of the brain involved. There is essential and mutual interaction between the relevant neuronal levels. Edelman and Tononi, for example, emphasize the global nature of conscious states and it is reasonable to interpret this as the view that conscious states are composed of both the higher and lower order states. They refer to what they call the “dynamic core” as generally “spatially distributed and thus cannot be localized to a single place in the brain.” (Edelman and Tononi, 2000: 146) It seems to me that their description of the neural correlates of consciousness fits more naturally with the WIV version of the HOT theory. (4) With respect to Levine’s objection, then, it is at best misleading to treat the lower and higher level “parts” of a conscious state as potentially “bifurcated.” While the standard AHOT theory might have a problem with respect to possible misrepresentation, it seems to me that the WIV is better able to handle this objection. This is because the “two” levels cannot really come apart in the way the Levine describes in his thought experiment. Levine’s options three and four, therefore, seem open to the HOT theorist and particularly open to a defender of the WIV.

3. Conclusion

In closing, then, the AHOT theory of consciousness is consistent with animal consciousness. Those who think otherwise, such as Peter Carruthers, typically build too much into the HOTs which render mental states conscious. Moreover, the AHOT theory remains preferable to the DHOT alternative. Secondly, the charge that the HOT theory cannot handle cases of misrepresentation can be answered. Authors such as Levine and Neander do not properly explore the options and resources available to the HOT theorist, and they hastily conclude that the HOT theorist must choose between arbitrary or clearly untenable alternatives.

NOTES:

1. No doubt that part of the problem here is terminological, as is often the case in the literature
on consciousness. One can, I suppose, *speak* of non-conscious ‘sufferings,’ ‘feelings,’ and ‘desire frustrations,’ and we are perhaps all entitled to use our own terminology to some extent. However, it seems to me that there is a point where using terms in this way becomes more of a provocative attempt to redefine them and can even simply add to the terminological confusion. It is most important, however, to keep our sights set on the substantial disagreement about whether or not animals have phenomenally conscious mental states.

2. For another more detailed criticism of Carruthers’ recent moral argument, see McKinnon, 2002.

3. Another important feature of Carruthers’ account is that some perceptual contents acquire a dual analog content, one purely first-order (e.g. ‘green’) and the other higher-order (e.g. ‘experience of green’). For Carruthers, it is the higher-order contents which confer subjectivity on the perceptual state; conscious experiences (percepts) are made conscious by virtue of their being available to the subject’s HOT forming module. See Carruthers (this volume) for more details.

4. Peter Carruthers has informed me (via email correspondence) that he is an “actualist” about introspection. Perhaps this is his best option for various reasons, though it may sound a bit surprising at first. However, I now wonder (a) why so little was said in his book about the structure of introspection, and, more importantly, (b) whether or not this detracts from his initial motivation for the dispositional account of first-order consciousness, especially when combined with the questions raised earlier about his “cognitive overload” argument against AHOT theory.

5. It actually seems to me that such tests for ‘other-attributing’ thoughts in the cognitive ethology and theory of mind literature are really aimed at determining whether or not animals or infants can have conscious HOTs directed at another’s mental state. Children are often even asked to verbalize their attitudes toward another’s beliefs or perceptions, and animals seem to be tested for behavioral signs that they are consciously thinking about another’s beliefs, e.g. in cases of deception. In my view, even if the evidence suggests that these subjects fail such tests, it causes no problem for the HOT theory since the theory certainly allows for the presence of conscious states in the absence of (either self-attributing or other-attributing) conscious HOTs. This is also one very brief line of reply to Seager (this volume).

6. It is worth noting that Rosenthal defends Levine’s “option 1” in several places. See, for example, Levine (2001: 190, note 24) for one place where Rosenthal says that if the HOT occurs without the target state the resulting conscious state might just be subjectively indistinguishable from one in which both occur. Once again, I find this option implausible partly because, as Levine says, “...doesn’t this give the game away?...then conscious experience is not in the end a matter of a relation between two (non-conscious) states.” (Levine, 2001: 190). Second, it seems to me that since the HOT is nonconscious, there would not be a conscious state anyway unless
there is also the accompanying lower-order state. Thanks to Alex Byrne for calling my attention to this point.
REFERENCES:
Carruthers, P. (this volume). HOP over FOR, HOT theory.


Rowlands, M. (this volume). Varieties of higher-order theory.


