



**Review of Peter Carruthers' *Consciousness: Essays From a Higher-Order Perspective***

Rocco J. Gennaro  
Department of Philosophy  
Indiana State University  
Root Hall A-138F  
Terre Haute, IN 47809  
USA  
rocco@indstate.edu

**PSYCHE 12 (4), August 2006**

**REVIEW OF:** *Consciousness: Essays From a Higher-Order Perspective*, Peter Carruthers, 2005. Oxford: Clarendon Press. 247pp. ISBN: 0199277354.

This is a fine and important collection of eleven recently published essays by Peter Carruthers, a leading figure in contemporary philosophy of mind. The book contains a very helpful introduction that provides a nice overview of Carruthers' basic views and orients the reader to the key issues. The introduction also presents a brief summary of the eleven chapters that comprise the remainder of the book. Only three of the essays initially appeared prior to Carruthers' important 2000 book *Phenomenal Consciousness: A Naturalistic Theory*. One of these is significantly rewritten: an essay entitled "Natural Theories of Consciousness," which will be familiar to readers of *Psyche* since it was the subject of a symposium in volumes 4-6. There is also one entirely new essay entitled "Dual-Content Theory: the Explanatory Advantages" (Chapter 6). Thus, there is much to learn about Carruthers' theory of consciousness even for those very familiar with his 2000 book. Indeed, one might think of this collection as elaborating upon Carruthers' views as found in his earlier book. In the relatively short period of a few years, Carruthers has managed to publish an impressive number of quality essays, which comprise most of this anthology. It is very convenient to have these nicely written essays all in one place. For those unfamiliar with *Phenomenal Consciousness*, this book is still readable on its own and contains many helpful summaries of Carruthers' earlier work.

In addition to the introduction (chapter 1), it is useful to think of *Consciousness* as divided into three parts. Part One contains chapters 2-6 and largely involves further development of Carruthers' so-called "dispositional higher-order thought (HOT) theory

of consciousness,” which he now prefers to call “dual-content theory.” There is also a chapter (chapter 2) about what it would take for phenomenal consciousness to be *reductively* explained (“Reductive Explanation and the ‘Explanatory Gap’”) in which Carruthers argues against those who claim that consciousness cannot be explained and must remain a mystery. Chapters 3-5 are “Natural Theories of Consciousness,” “HOP over FOR, HOT Theory,” and “Phenomenal Concepts and Higher-Order Experiences.” Part Two contains chapters 7 and 8 that focus on conscious *thinking* and conscious propositional thought, as opposed to conscious *experience*. In these chapters, entitled “Conscious Thinking: Language or Elimination?” and “Conscious Experience versus Conscious Thought,” Carruthers basically shows how his theory can be extended to account for conscious thinking. These essays also explore “how natural language might be both constitutive of, and necessary to the existence of, conscious propositional thought-contents,” and Carruthers endorses a “form of eliminativism about thought *modes* (believing versus desiring versus supposing, etc.)...” (p. 16). Part Three includes chapters 9-12 which focus on the mental lives of non-human animals, the most controversial aspect of Carruthers’ overall account, as we shall see. The chapter titles are “Sympathy and Subjectivity,” “Suffering without Subjectivity,” “Why the Question of Animal Consciousness Might Not Matter Very Much,” and “On Being Simple Minded.”

In what follows, I will focus on four central and interrelated themes found primarily in Parts One and Three. It is, of course, impossible to do justice to all of the subtleties in every chapter. However, I will attempt to explain and critically evaluate many of the central ideas, sometimes contrasting Carruthers’ views with my own.

## 1. Higher-Order (HO) versus First-Order (FO) Representationalism

Representationalism, in general, is the view that the phenomenal properties of experience (i.e. “qualia”) can be explained in terms of the experiences’ representational properties. To place Carruthers’ theory of consciousness in its proper context, it is useful to keep in mind that there are numerous versions of representational theory, but the most basic difference is between first-order (FO) theories and higher-order (HO) theories. FO theories attempt to explain conscious experience in terms of world-directed (or first-order) intentional states, whereas HO theory holds that some kind higher-order state is needed to render a state conscious. There are various kinds of HO theory with the most common division between higher-order *thought* (HOT) theories and higher-order *perception* (HOP) theories. HOT theorists, such as David Rosenthal, think it is better to understand the HO state as a thought of some kind. HOTs are treated as *cognitive* states involving some kind of conceptual component. HOP theorists, such as William Lycan, urge that the HO state is a *perceptual* or *experiential* state of some kind which does not require the kind of conceptual content invoked by HOT theorists. Much of Part One is devoted to defending Carruthers’ version of HO theory while criticizing FO views, such as those put forth Michael Tye and Fred Dretske.

Carruthers argues in various places that FO accounts cannot properly distinguish between non-conscious and (phenomenally) conscious mental states, or between what he calls “worldly subjectivity” and “experiential subjectivity.” He explains that “any first-order perceptual state will be, in a sense, subjective. That is, it will present a subjective take on the organism’s environment...But phenomenal consciousness surely involves a

much richer form of subjectivity than this...and [has] a distinctive *feel* or phenomenology” (p. 70). And “mental states are phenomenally conscious when it is like something to undergo them...” (p. 3). Carruthers offers several distinct and elaborate arguments against the plausibility of FO theory, and it is impossible to discuss them all here. In his previously unpublished chapter six, Carruthers argues that FO theories are “incapable of explaining the distinctive, puzzling, to-be-explained features of phenomenal consciousness” (p. 100). Carruthers lays out six desiderata for a successful (reductive) theory of consciousness, including why phenomenally conscious states have a subjective aspect and why there seems to be an explanatory gap between the phenomenal and physical/functional. He explains in great detail why FO theory fails to explain these desiderata whereas his version of HO theory can do so. For example, he argues that FO theories are not explanatory, in part, because it remains mysterious just how the intentional contents in question can be transformed from states with mere worldly subjectivity to states that are phenomenally conscious. For instance, on Tye’s view, part of what is supposed to do the work is that contents of phenomenal states are available to make an impact on one’s beliefs and/or desires. This is a functional notion but, Carruthers urges, “it just isn’t clear why this sort of availability should give rise to states that it is like something to undergo...” (p. 103). Throughout the book, Carruthers relies heavily on the so-called “two-systems theory of vision” (following Milner and Goodale) whereby the perceptual states produced by the (human) ventral/temporal system are phenomenally conscious ones whereas those produced by the dorsal/parietal “how-to” action system aren’t (see pp. 44, 72-3, 98-9, 199-201). Although an FO theorist could also accept this theory of vision, it would still remain unclear just *why* the ventral/temporal states are conscious since they would also merely produce intentional contents which represent distal properties of the environment and, hence, could at best produce worldly subjectivity.

As a defender of a HO theory myself (Gennaro 1996, 2005), I tend to agree with much of what Carruthers says in his rejection of FO theory. I will now turn to his more positive view.

## 2. Dual-Content Theory

Carruthers had previously called his theory of consciousness the “dispositional HOT (higher-order thought) theory,” but he now refers to it as the “dual-content theory.” He believes that it is better to think of the HOTs as *dispositional* states instead of the standard view that the HOTs are *actual*. The basic idea is that the conscious status of an experience is due to its *availability* to higher-order thought. There are numerous places where dual-content theory is characterized in more detail (e.g. pp. 8-15, 55-60, and 64-66). For example, “phenomenal consciousness consists in a certain sort of intentional content (‘analog’ or fine-grained) that is held in a special-purpose functionally individuated memory store in such a way as to be available to a faculty of higher-order thought” (p. 8). (Intentional contents are *analog* when they have a finer grain than any concepts that the subject can possess and recall.) Thus, some first-order perceptual contents are available to a higher-order “theory of mind mechanism,” which transforms those representational contents into conscious contents (though no actual HOT occurs). According to Carruthers, some perceptual states acquire a dual intentional content; for example, a conscious experience of red not only has the first-order content ‘red,’ but also

has the higher-order content ‘seems red’ or ‘experience of red.’ Carruthers makes frequent use of so-called “consumer semantics” in order to fill out his theory of phenomenal consciousness. The basic idea here is that the content of a mental state depends, in part, on the powers of the organism which “consume” that state, e.g. the kinds of inferences which the organism can make when it is in that state. It should also be noted that (perhaps somewhat curiously) Carruthers understands his “dispositional HOT theory” to be a form of HOP theory (see especially chapter four). He explains that it is because dual content theory “proposes a set of higher-order analog—or ‘experiential’—states, which represent the existence and content of our first-order perceptual states, that the theory deserves the title of ‘higher-order *perception*’ theory, despite the absence of any postulated *organs* of higher-order perception” (p. 64).

I have argued elsewhere against Carruthers version of HO theory and in favor of a position closer to actualist HOT theory (Gennaro 2004). Without going into great detail, however, there are at least several standard lines of criticism: First, Carruthers’ main objection to the actualist HOT theory is based on what he calls “cognitive overload” (p. 54). The objection is that actualist HOTs would take up too much “cognitive space” (i.e. neural space”) given the immense amount that we can experience consciously at one time. Carruthers rejects the reply that our conscious experience is not as rich and complex as it might seem, and thus believes that dual content theory fares better on this point. Since the HOTs are not actual, less cognitive space is needed and so the objection is avoided. However, it is unclear that this is really the case regardless of how complex conscious states are. As Carruthers makes clear, dual content theory still posits the presence of *actual* brain structures to fill out his theory (e.g. theory of mind mechanism, special memory store). More generally, dispositional states also require similar brain structure because something categorical (i.e. actual) must underlie any disposition. Moreover, it is not clear why an actual state with two contents takes up less cognitive space than two states with one representing the other.

Second, many will wonder just how dual content theory explains phenomenal consciousness. For one thing, it is difficult to understand how a dispositional HOT can render, say, a perceptual state *actually* conscious. To be fair, Carruthers is well aware of this objection and attempts to address it (e.g. pp. 55-60). He leans heavily on consumer semantics here and uses several experimental results (e.g. prosthetic vision, spatially-inverting lenses) in an attempt to show that “changes in consumer systems can transform perceptual contents...” (p. 56). No doubt many will remain dissatisfied for various reasons, even many who are otherwise sympathetic to his overall approach. Moreover, one might wonder if dual content theory is vulnerable to (almost) the same objection raised by Carruthers against FO theory that we saw earlier. On both accounts, it is still difficult to see how the functional/dispositional aspects of the respective theories can yield actual conscious states. It is one thing to show that conscious intentional *contents* can be transformed or altered due to such considerations, but arguably another to claim that they can render an otherwise non-conscious state *conscious*.

Third, one might also therefore wonder why Carruthers’ theory is a theory of *consciousness* as opposed to a theory of *content*, especially given the fact that he relies so heavily on consumer semantics to fill out his theory. As Carruthers himself says, “according to all forms of consumer semantics (including teleosemantics and various

forms of functional and inferential role semantics) the intentional *content* of a state depends, at least in part, on what the down-stream consumer systems that can make use of that state are disposed to do with it” (p. 56, emphasis added). That is, perhaps consumer semantics can explain why a state has the content it has, but that is not quite the same as explaining why the state is conscious in the first place. (See Rosenthal 2004: 24-30, and Jehle & Kriegel 2006 for related criticisms.)

Nonetheless, one very interesting aspect of Carruthers’ view is that conscious states, in some sense, represent themselves. The notion that self-representation is an essential aspect of conscious states has a long history, though there are many different versions of such a theory, ranging from phenomenological (non-reductive) accounts to more recent naturalistic variations of HO theory. I am somewhat sympathetic to this approach (Gennaro 2006; see also Kriegel and Williford 2006 for numerous papers on this topic). That Carruthers sees dual-content theory as a kind of self-representationalist theory of consciousness is clear in many chapters, but consider the claim, for example, that “phenomenally conscious [experiences]...come to present themselves to us, as well as presenting properties of the world (or of the body) represented.” (p. 107; see also pp. 65-6 and chapter eight). We can thus think of a conscious state, on Carruthers’ account, as having two contents, one first-order and the other higher-order. An interesting aspect of all this is that his view, if tenable, can also be used as a reply to a very common objection to standard actualist HOT and HOP theory; namely, that they cannot account for what happens when (or if) the HO state *misrepresents* the lower-order (LO) state (see Gennaro 2004 for some discussion and further references). As Carruthers says, “it should be obvious why there can be no question of our higher-order analog contents getting out of line with their first-order counterparts, on this account...this is because the higher-order experience *seems [red]* is parasitic on the content of the first-order experience *[red]*, being formed from it by virtue of the latter’s availability to a ‘theory of mind’ system” (p. 96). Thus, if this version of HO theory can guarantee a match between the HO and LO content, then this powerful objection is defused.

### 3. Reductive Explanation and Phenomenal Concepts

Two importantly related themes in Carruthers’ work are the ideas that consciousness can indeed be reductively explained and that there are so-called “phenomenal concepts” (Loar 1997). Carruthers’ optimism about the prospects of explaining consciousness in a reductive fashion is difficult not to notice: “It turns out that the ‘hard problem’ isn’t really so very hard after all” (p. 60). There are those seek to reduce consciousness more directly in neural or physical terms, but I agree with Carruthers that doing so tries “to leap over too many explanatory levels at once.” (p. 6). A major part of Carruthers’ strategy to explain why there at least *seems* to be an “explanatory gap” between the mental and physical relies on the notion that are “phenomenal” (or “recognitional”) concepts. Purely recognitional concepts are those “we either have, or can form...that lack any conceptual connections with other concepts of ours, whether physical, functional, or intentional. I can, as it were, just recognize a given type of experience as *this* each time it occurs, where my concept *this* lacks any conceptual connections with any other concepts of mine – even the concept *experience*” (p. 67). Part of the motivation behind Carruthers’ acceptance of phenomenal concepts is to counter various well-known “zombie” and “conceivability” thought-experiments used against reductive materialism. Thus, for

example, in chapter two, Carruthers argues, among other things, that any alleged explanatory gap can be explained away because if we possess purely recognitional concepts of the form ‘*This* type of experience,’ we will still always be able to have that thought while, at the same time, conceiving of the absence of any corresponding physical or intentional property. However, there still may only be the two different *concepts* referring to one and the same *property*. On the one side, we are dealing with scientific third-person concepts and, on the other, we are employing phenomenal concepts. We are, perhaps, simply not in a position to understand completely the connection between the two.

It is worth noting that recognitional concepts also play a crucial role in many of Carruthers’ arguments favoring dual content theory over other theories. For example, he argues at length that neither FO nor actualist HOT theory can explain the presence of recognitional concepts (in especially chapters 4-6). Moreover, Carruthers argues against the existence of “qualia,” understood as intrinsic (non-relational, non-intentional) properties of experience, partly on the basis of his acceptance of recognitional concepts.

#### 4. Animal Mentality and Consciousness

The most controversial aspect of Carruthers’ views concerns his position on animal consciousness. I have had my say elsewhere on Carruthers’ contention that animal consciousness is very unlikely given the truth of some form of HO theory (Gennaro 1993, 1996, 2004). This is not the place to expand greatly on those replies. However, a brief summary and a few additional thoughts are in order. Following up on earlier work, there is much discussion of animal consciousness in Part Three of this book, but, before addressing those chapters explicitly, I first pause to reiterate my counter-reply to a previous reply from Carruthers’ (see pp. 49-53).

One of the most common objections to HOT theory comes from the concern that it rules out or, at least, renders unlikely the possibility of animal (and even infant) consciousness. This objection is frequently raised in the literature by those who reject HO (especially HOT) theory. I have argued in response that the HOT need not be as sophisticated as it might seem. 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. Most HO theorists, therefore, are concerned to show that HO theory does not rule out animal consciousness. Carruthers, however, accepts the consequence that the mental states of many animals lack phenomenal consciousness.

Carruthers presents the following summary statement of my position as follows: “[i]n order for [mental state] M to count as conscious, one doesn’t 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” (p. 49). 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 color experiences as opposed to just colors. 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... (p. 50)

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” (p. 50). 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 like ‘looking red’ or ‘seeing red’? Is it so implausible to ascribe *these* concepts to most animals? I don’t think so. Animals need not have the concept of ‘the *experience* of red’ as opposed to ‘seeing or looking red.’ “I am now seeing red” is a perfectly good HOT. Similarly, even if animals do not have HOTs containing the concept ‘experience’ in any sophisticated sense of the term, 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? 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.

Moreover, as we have seen, Carruthers champions the view that there are purely recognitional concepts of *experience*. Recall that the relevant thoughts do not even have to contain the concept ‘experience’: they are those concepts “we either have, or can form...that lack any conceptual connections with other concepts of ours, whether physical, functional, or intentional. I can, as it were, just recognize a given type of experience as *this* each time it occurs, where my concept *this* lacks any conceptual connections with any other concepts of mine – even the concept *experience*” (p. 67). One might therefore wonder why *we* can have such stripped down concepts but animals can’t. There is arguably some tension in Carruthers’ view here. Why do animals need to have the concept ‘experience’ in their HOTs (making it less likely that they are conscious creatures) but yet we don’t even need to have such sophisticated HOTs? Indeed, the presence of something like recognitional concepts seems to be precisely what Carruthers doesn’t, but should, allow in response to option one.

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” (p. 50) But I have argued (in Gennaro, 1996: 95-101) that the difference between the HOT and HOE [= HOP] models is greatly exaggerated. Contrary to what Carruthers says (p. 50, fn. 18), I have never argued that there is *no* real difference between HOT and HOP theory. However, part of my objection rests on my conceptualist tendencies, so that I am skeptical that there are even HOPs (or any perceptions, for that matter) with analog content, let alone entirely non-conceptual content. Others have also questioned the traditional division of HO theories for different reasons (Van Gulick, 2000), and some have 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. If anything, it seems to me that Carruthers' view should make it *more* likely that animals are phenomenally conscious. HOP theory is normally viewed as having less of a problem in accounting for animal consciousness precisely because the HO state is (at least) partly non-conceptual and more like a perceptual state. Thus, Carruthers holds a form of HO theory which is normally even friendlier to animal consciousness. And, of course, Carruthers himself blurs the distinction between the HOP and HOT models by arguing that his dispositional HOT theory is a form of HOP theory. So it is difficult to see why any move in that direction would be so problematic for my view either.

Moreover, in previous 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 that 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 (e.g. various emotions). While Carruthers is clearly very knowledgeable about brain science and evolution, his failure to put the above disagreement in context is significant, because the *cumulative* effect of such strong inductive evidence in favor of animal consciousness is lost.

Now why exactly does Carruthers think that most animals don't have HOTs? A primary reason 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. Once again, however, Carruthers builds a great deal into having such a theory and then explicitly ties it to the capacity to have HOTs at all. He cites experimental work by Povinelli and others suggesting that very young children and even some chimps lack the concepts of 'appearing' or 'seeming' or 'perception' (as subjective states of the perceiver), which can be taken as necessary to entertain higher-order thoughts about experiences. Such experiments are often designed to determine if chimps take notice of whether or not the experimenter is looking at or away from something (such as food). In line with many 'theory-theorists,' Carruthers holds that animals with HOTs should be able to have thoughts about the mental states of *other creatures* as, for example, we might expect to find when (or if) animals engage in deceptive behavior. (He is also not convinced by evidence for animal metacognition based on so-called "uncertainty monitoring" tests; see pp. 206-214.)

I strongly disagree with the notion that one should conclude from such evidence that (most) animals don't have HOTs. I'll make a few brief points here: *First*, it is not at all clear that so much should be read into the failure of animals in such experiments. For one thing, these are obviously not the natural conditions or environments of the animals in question. Perhaps failure can be explained in such situations because they don't typically arise in their native environment. My understanding is that many primates, at the least, do much better in similar tests when performed in more natural settings (as was persuasively shown at the 2006 Society for Philosophy and Psychology meeting in a presentation by Laurie Santos entitled "The Evolution of Mind Reading: Insights from Non-Human Primates"). But, in any case, there are some questionable interpretations of various experimental results. *Second*, even if some or most animals cannot, say, engage in deceptive behavior and so arguably 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. After all, self-directed HOTs are all that are required for conscious states, according to HOT theory. Thus, I agree with Ridge (2001) 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. To be sure, this is a major issue in its own right; namely, to what extent the HOT model requires "mind-reading" of others as opposed to mere "self-monitoring." Nonetheless, the former seems far more sophisticated and is not necessary for there simply to be conscious mental states, such as pains. *Third*, it even seems to me that some tests for 'other-attributing' thoughts in the cognitive ethology and theory of mind literature are often aimed at determining whether or not animals or infants can have *conscious* HOTs directed at another's mental state. For example, young children are 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. But even if the evidence suggests that these subjects fail such tests, it causes no problem for HOT theory since it certainly allows for the presence of conscious states in the absence of (either self-attributing or other-attributing) *conscious* HOTs. As is well-known, the HOT theory only requires *non-conscious* HOTs to render a first-order state conscious. If the HOT is itself conscious, then one is in a more sophisticated *introspective* state, which is not necessary for having a first-order conscious state.

Linking his discussion of animal consciousness 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. But Carruthers has changed his mind. In Chapters 9 and 10, Carruthers is concerned to show that animals can be objects of sympathy and moral concern because the "most basic form of mental...harm lies in the existence of thwarted agency, or thwarted desire, rather than in anything phenomenological" (p. 157). Indeed, according to Carruthers, even suffering, grief, and disappointment, can occur in the absence of phenomenal consciousness because organisms, who are still capable of having (non-conscious) mental states, can find themselves in a situation such that there is "the co-activation within a creature's practical reasoning system of a first-order desire together with the first-order belief that the state of affairs that is the object of the desire doesn't obtain" (p. 177). Suppose that an animal currently wants to drink but believes that it isn't presently drinking. And animals can be still averse to non-conscious pains in the sense that they take steps to avoid being in that state. Carruthers interestingly argues in various places (especially chapter 12) for the view that it is perfectly reasonable to attribute all kinds of (non-conscious) mental states to animals, even to ants and bees.

I find Carruthers' arguments for his current moral stance unconvincing. For example, he asks us to imagine 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." (p. 172). 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" (p. 173). In essence, Carruthers is attempting to separate desire frustration from

consciousness in order to make room for the idea that non-conscious animals can indeed be the objects of sympathy and moral concern, contrary to his previously held position.

I am puzzled by Carruthers' arguments and view here for several reasons. First, even the hypothetical Phenumb begins as a conscious agent. It seems to me that desire frustration is an even more sophisticated 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, even if we can imagine the possibility of some Spock-like character only able to have such purely intellectual frustrations (as Carruthers suggests on p. 172 fn. 15), it does not follow that such frustrations would be entirely non-phenomenal. Carruthers is curiously comparing (what he takes to be) non-conscious animals to a very sophisticated intellectual hypothetical character. Third, 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." 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. Thus, in the end, I don't believe that Carruthers' current moral stance is any more tenable than his previously held view. No doubt that part of the problem might be 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, there is a point where using terms in this way becomes more of a provocative attempt to redefine them and simply adds to the terminological confusion. It is most important, though, to keep our sights set on the issue of whether or not animals have phenomenally conscious mental states.

Carruthers might at this point accuse me of implicitly holding a Searlean position such that each non-conscious mental state is either actually or potentially conscious. But I hold no such position and there is a middle view ignored by Carruthers; namely, that some kinds of mental states (e.g. pains, frustrations, sufferings) can only be had by a conscious organism (see Gennaro 1996, chapters 1 and 5). That is, from the fact that *we* have non-conscious pains and feelings, it doesn't follow that there are (or could be) organism with *all* non-conscious pains and feelings. Another way to think about this is by comparing certain primitive animal behavior to some current day robots. If we are convinced that a robot is utterly non-conscious, we may still (rightly, I think) attribute to it beliefs, desires, and perhaps even perceptions, if its behavior is complex enough. However, it is not clear that the same would or should go for pains, sufferings, frustrations, feelings, and disappointments. These are arguably parasitic on the *creature* (or system) being conscious in the first place. It seems to me that we wouldn't (and shouldn't) ever say that a robot is suffering or in pain unless we were convinced that it is capable of otherwise having phenomenally conscious states. Indeed, we are often inclined to think in terms of the entire organism as a conscious agent when attributing such states as frustrations or disappointments: "I am frustrated," "The dog is suffering," and "My sister is very disappointed." But this is not to endorse either the stronger Cartesian view or even the Searlean position and is consistent with holding that any individual mental state in a conscious creature can be non-conscious. The key difference, I think, lies in the fact that beliefs and desires are best understood purely as dispositions to behave in certain

ways, and thus are more reasonably attributed to utterly non-conscious robots or even to, say, some non-conscious insects. Carruthers, however, does raise the very important issue of the relationship between intentional states (such as beliefs and desires) and consciousness. And he clearly also treats desires and beliefs as somewhat more primitive mental states, which can even be attributed to ants and bees (chapter 12). In these cases, the behavior in question is at least arguably complex enough to warrant such mental ascriptions. It is at least a notch up from the purely inflexible fixed-action behavior patterns found in some primitive insects.

It is also unlikely that the common belief that many animals have conscious states (which in turn cause their behavior) can be so easily explained away as an anthropomorphic process of “imaginatively projecting” what it is for us to have certain mental states onto various animals (as Carruthers tries to do in response to Robert Lurz on pp. 198-200). Carruthers insists that we are under an illusion in thinking that phenomenal consciousness is needed to explain the cause of animal (and even our) behavior. He calls it the “in virtue of illusion” whereby we mistakenly think that it is in virtue of the phenomenally conscious properties of experience that we behave the way we do. If someone picks out a tomato by its color, one will normally have a phenomenally conscious experience of red but it doesn’t follow that the phenomenal property *causes* the behavior in question. For Carruthers, then, although a human phenomenally conscious *state* does cause the behavior in question, it is the first-order (not higher-order) *content* of a conscious state that does virtually all of the causal work. Carruthers contends that a similar explanation can be given for a variety of animal behaviors except that they are not phenomenally conscious in the first place (i.e. there’s no higher-order content to the state at all). This move by Carruthers strikes me as highly implausible and certainly does not seem to describe what I am thinking when I attribute conscious mental states to animals. I agree with Lurz that the central initial reason for believing that animals have conscious mental states has more to do with the fact that their rather complex behavior is best explained and predicted by attributing such folk psychological notions to them. We *could* be wrong of course and under some massive illusion here, but I am not convinced that we are. In addition, Carruthers’ view seems undermined by the similarity in human and animal brain structure.

As Carruthers recognizes, he is thus dangerously close to embracing some form of epiphenomenalism whereby conscious states (or at least the property of consciousness) have no causal impact on one’s behavior (see e.g. pp. 174, 183, 186-7, 195, 198, 204-6). But, in some places, he doesn’t wish to go that far saying that “phenomenal consciousness might be *almost* epiphenomenal in its functioning within human cognition...” (p. 195). Carruthers has much to say about how animal behavior can be explained without appeal to consciousness, but one begins to wonder just why we are forced to attribute consciousness to other humans. He relies on the aforementioned two visual systems hypothesis and cites familiar evidence from blindsight cases to show how some behavior can occur non-consciously (e.g. pp. 204-6). However, he overlooks the fact that blindsight patients do not voluntarily act toward the objects in their blind fields. They only act in response to forced guessing in response to the examiner. So it doesn’t seem to me that blindsight cases support Carruthers’ view, at least to the extent that animals behave voluntarily or “on their own” towards objects of perception. We do indeed seem to behave, at least sometimes, in virtue of the consciousness of our

perceptions. Furthermore, I am not aware of there being any equally clear “two-system” analogy for all of the other sensory modalities, such as touch, hearing and smell. For example, an unexpected sound behind an animal may cause it to flee. The animal is not being forced to guess whether or not something is behind it. The best explanation still seems to be that it fled because it consciously heard the sound, at least for animals with complex enough behavior and comparative brain areas (and the same might go for certain smells). How could non-conscious visual experiences (and smells and sounds) cause a seeing-eye dog to help its owner across a busy street? After all, the owner needs the dog precisely because she has lost her *conscious* vision. Similarly, navigating through the world by means of blindsight is obviously not good enough, not to mention very dangerous. One can only imagine how much worse off she would be if she also lost her conscious sense of touch or smell or hearing. But if Carruthers is right, having a non-conscious dog is somehow able to help a blind person to a great extent. This is very puzzling in my view. Moreover, if we have two visual systems (even if one is entirely non-conscious and the other conscious), then it is surely reasonable to hold that animals, whose brain structures are similar, also have the conscious system. Carruthers might again insist that those animals don’t have a “HOT faculty.” But then why would evolution have produced two visual systems in so many animals if they are both non-conscious?

Despite Carruthers’ insistence that he has “no axe to grind” (e.g. p. 181), it is difficult not to notice the elaborate and creative attempts made to explain away every bit of evidence for animal consciousness (and HOTs) as misleading, mistaken, or illusory while any piece of evidence suggesting the presence of (non-conscious) intentional states is interpreted in the most favorable light. To be sure, Carruthers often presents very interesting reasons and important arguments in support of his position. But it is a very puzzling strategy and combination of views, in my opinion. Moreover, it is rather surprising that Carruthers could say that there “is no radical Cartesian divide here, between genuinely minded humans and mere mindless automatons...” (p. 204). This may be true, of course, if we were talking only about non-conscious mental states reaching far down the evolutionary chain, but, of course, the “great Cartesian divide” has much more to do with any radical difference between us and animals in terms of *consciousness*. That is certainly the real Cartesian divide; after all, Descartes didn’t even believe in non-conscious intentional states. It is clear to me that humans would (and should) treat animals differently if we were convinced that they were not conscious, contrary to Carruthers’ claim that “very little of significance for comparative psychology need follow from the fact that phenomenal consciousness is denied to many non-human animals by higher-order thought theories of consciousness.” (p. 204) Finally, given his fairly radical views about animal consciousness, one wonders just what entitles Carruthers to appeal to “common sense” and “authority” in a number of other places when it suits his purposes (e.g. pp. 216-7). Why should anyone accept what “most people” think about, say, ascribing beliefs and desires to animals from someone who rejects similar logic regarding animal consciousness?

Overall, however, this is a very important volume that is rich with interesting interdisciplinary discussion and packed with thought provoking arguments. It is a must-read for anyone interested in philosophy of mind and consciousness research. Even when one disagrees with Carruthers, there is much to learn from his writings.

## References

- Carruthers, P. (2000). *Phenomenal Consciousness*. Cambridge, MA: Cambridge University Press.
- Gennaro, R.J. (1993). Brute experience and the higher-order thought theory of consciousness. *Philosophical Papers*, 22, 51-69.
- Gennaro, R.J. (1996). *Consciousness and Self-consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. Amsterdam & Philadelphia: John Benjamins.
- Gennaro, R.J. (2004). Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In Gennaro 2004a.
- Gennaro, R.J. (ed. 2004a). *Higher-Order Theories of Consciousness: An Anthology*. Amsterdam & Philadelphia: John Benjamins.
- Gennaro, R.J. (2005). The HOT theory of consciousness: between a rock and a hard place? *Journal of Consciousness Studies*, 12 (2), 3-21.
- Gennaro, R.J. (2006). Between pure self-referentialism and the extrinsic HOT theory of consciousness. In Kriegel and Williford 2006.
- Guzeldere, G. (1995). Is consciousness the perception of what passes in one's own mind? In T. Metzinger. (Ed.), *Conscious Experience*. Paderbom: Ferdinand Schöningh.
- Jehle, D. & Kriegel, U (2006). An argument against dispositionalist HOT theory. *Philosophical Psychology*, forthcoming.
- Kriegel, U. & Williford, K. (eds. 2006). *Self-Representational Approaches to Consciousness*. Cambridge, MA: MIT Press.
- Loar, B. (1997). Phenomenal states. In N. Block, O. Flanagan, and G. Guzeldere (eds). *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Ridge, M. (2001). Taking solipsism seriously: nonhuman animals and meta-cognitive theories of consciousness. *Philosophical Studies*, 103, 315-340.
- Rosenthal, D. M. (2004). Varieties of higher-order theory. In Gennaro 2004a.
- Van Gulick, R. (2000). Inward and upward: reflection, introspection and self-awareness. *Philosophical Topics*, 28, 275-305.