

The HOT Theory of Consciousness: Between a Rock and a Hard Place?

Rocco J. Gennaro

I. Introduction

The so-called “higher-order thought” (HOT) theory of consciousness says that what makes a mental state conscious is the presence of a suitable higher-order thought directed at it (Rosenthal 1986, 1990, 1993, 2002, 2004; Gennaro 1996, 2004).¹ The HOT theory has been or could be attacked from two apparently opposite directions. On the one hand, there is what Stubenberg (1998) has called “the problem of the rock” which, if successful, would show that the HOT theory proves too much. On the other hand, it might also be alleged that the HOT theory does not or cannot address the so-called “hard problem” of phenomenal consciousness. If so, then the HOT theory would prove too little. We might say, then, that the HOT theory is arguably between a rock and a hard place. In this paper, I critically examine these objections and defend the HOT theory against them. In doing so, I hope to show that the HOT theory, or at least some version of it, neither proves too little nor too much, but is just right. I also show that these two objections are really just two sides of the same coin, and that the HOT theory is immune from David Chalmers’ (1995, 1996) criticisms of other attempted reductionist accounts of consciousness.

II. The Problem of the Rock

Following Stubenberg (1998), I will call the following classic objection from Alvin Goldman to HO theories of consciousness “the problem of the rock”:

the idea here is puzzling. How could possession of a meta-state confer subjectivity or feeling on a lower-order state that did not otherwise possess it? Why would being an intentional object or referent of a meta-state confer consciousness on a first-order state? A rock does not become conscious when someone has a belief about it. Why should a first-order psychological state become conscious simply by having a belief about it?” (Goldman 1993, p. 366)

Clearly, this objection could be devastating. If successful, then HO theories would be guilty of proving too much (i.e. they would make too many things conscious) and would thereby be reduced to absurdity. Thus, a HO theorist must block the generalization to rocks and the like

¹ Some of what I argue applies both to the HOT theory and so-called “higher-order perception” (HOP) models (e.g. Lycan 1996). In those cases, I will use the more generic ‘HO’ abbreviation.

without sacrificing an informative analysis of consciousness. Two preliminary points must first be made on behalf of HO theories: (1) Every HO theorist acknowledges that the meta-state does not *always* confer consciousness on its target state. Thus, for example, the HOT must arise in a suitably unmediated way in order for it to confer consciousness. The HOT must not arise via inference from observing one's own behavior (Gennaro 1996, pp. 84-87). (2) The HO theorist might also object to Goldman's use of the term 'belief' instead of 'thought.' For example, one might argue that beliefs are best understood as dispositional states, and so are not up to the task of conferring consciousness on anything, including both rocks and mental states. Of course, it is not always easy to distinguish between an *occurrent* belief and a thought,² so perhaps the problem of the rock can simply be recast in terms of occurrent belief.

Nonetheless, Goldman's analogy is still suspect in at least two ways: (a) As Lycan observes, the difference between rocks and psychological states is simply that the latter "are themselves mental...It seems psychological states are called "conscious" states when we are [aware] of them, but nonpsychological things are not." (Lycan 1996, p. 24) Thus, in Goldman's analogy we do not first have a mental state which then becomes the object of a meta-state. Instead, there is a rock, clearly a "nonpsychological thing," which becomes the object of a mental state.³ (b) Goldman does not distinguish between conscious and unconscious meta-states in the way that HOT theorists do. The HOT theory says that what makes a first-order world-directed mental state conscious is the presence of an *unconscious* HOT directed at it, but, when the HOT is itself conscious, we then have an *introspective* state.⁴ So is the belief about the rock conscious or unconscious? It is difficult to make sense of the analogy in this context. On the other hand, perhaps Goldman and others can simply insist that the problem remains either way: Neither conscious nor unconscious thoughts about rocks make them conscious.

Thus, I still think that this objection needs to be taken seriously. I am also not satisfied with David Rosenthal's well-known early attempt to answer to this problem. He first distinguishes between transitive and intransitive consciousness: the former is the "conscious of" sense which is typically attributed to subjects, whereas the latter is the "is conscious" sense which is often also attributed to individual mental states (Rosenthal 1990, cf. Rosenthal 1993). He also defines an intrinsic property as "if something's having it does not consist, even in part, in that thing's bearing some relation to something else." (1990, pp. 21-22) He then replies to the problem of the rock as follows:

² But see Gennaro 1996, pp. 36-43 for one such attempt.

³ Cf. Byrne 1997, pp. 110-111.

⁴ See Rosenthal 1986, pp. 337-8.

Being transitively conscious of a mental state does in a sense make it intransitively conscious. But that is not because being conscious of a mental state causes that state to have the property of being intransitively conscious; rather, it is because a mental state's being intransitively conscious simply consists in one's being transitively conscious of it. The mistake here is to suppose that a state's being intransitively conscious is an intrinsic property of that state. If it were, then being intransitively conscious could not consist in one's being transitively conscious of being in that state unless being thus conscious induced a change in that state's intrinsic properties. This objection is at bottom just a disguised version of the doctrine that being intransitively conscious is an intrinsic property. (Rosenthal 1990, pp. 30-31)

First, it is not clear to me exactly why Rosenthal thinks that this objection is "just a disguised version of the doctrine that being intransitively conscious is an intrinsic property" of mental states. I can easily imagine someone very sympathetic to his view that HOTs are *extrinsic* to the states rendered conscious also being troubled by this problem. Indeed, Goldman's initial formulation of the problem seems quite explicitly to separate the meta-state from the lower-order state. But, once again, Goldman ultimately wants to know exactly *how* or *why* such meta-states can confer consciousness on their objects. As Byrne (1997, p. 110) notes, at worst the belief "...that intransitive consciousness is intrinsic is a plausible *consequence* of the objection, not the basis of it."

Second, Rosenthal accuses the proponents of this objection of believing, like Descartes, in the intrinsic nature of consciousness. But he also mistakenly seems to think it follows from this that consciousness would then be essential to all mental states and thus unanalyzable (or "simple").⁵ Rosenthal is so concerned to avoid the notion that consciousness is an intrinsic

⁵ See, for example, Rosenthal 1986, especially pp. 340-348; and Rosenthal 1990, pp. 22-24.

Intrinsic properties, as defined by Rosenthal, clearly need not be essential properties: my having dark hair is intrinsic to me but not essential. And conscious mental states can also have an informative and analyzable structure even the HOTs are intrinsic to them (see Gennaro 1996, pp. 21-30 for more on these and related points; cf. Schröder 2001, pp. 33-4). Indeed, there is no logical connection at all between such notions as 'intrinsicity,' 'extrinsicity,' and 'essentiality.'

property of any mental state that he overlooks the possibility that conscious mental states might consist of *both* the lower-order state and the conscious making HOT. I am very sympathetic to this “intrinsicist” version of the HOT theory and have defended it elsewhere (Gennaro 1996, chapter two; 2002, forthcoming). The idea here is that first-order conscious states are better understood as complexes of two parts (cf. Brentano 1874/1973).⁶ Notice that, on this view, the HOT still need not be essential to the very existence of the lower-order mental state and consciousness does not thereby become unanalyzable. But if Rosenthal adopts this view, then he has two problems: (a) The HOT (or “conscious making property”) would be intrinsic to the conscious state, which is clearly at odds with Rosenthal’s considered view, and (b) Rosenthal would then not be able simply to dismiss proponents of the problem of the rock as those opponents who just mistakenly treat consciousness as an intrinsic property of intransitively conscious mental states.

In any case, I think that Stubenberg is right when he says that “what motivates the proponent of the problem of the rock is the worry that the relation *being the target of a higher-order representation* is the wrong relation. The worry is fueled by one’s inability to comprehend how entering into this relation is supposed to promote an unconscious state to consciousness. Those who raise this objection aren’t just begging the question against Rosenthal by simply restating their intrinsicist creed.” (1998, p. 195) However, unlike Stubenberg, I do not think that the problem of the rock should cause one to give up on some version of a HO theory of consciousness.

So where do we go from here? It is first necessary to return to Lycan’s response to the problem. Although he did not go far enough, Lycan does take the first crucial step. We must first and foremost distinguish rocks and other nonpsychological things from the psychological states that HO theories are attempting to explain. HO theories must maintain that there is something not only special about the meta-state (as we will see in the next section), but also something special about the *object* of the meta-state, which, when combined in certain ways, result in a conscious mental state. The HO theorist must initially boldly answer the problem of the rock in this way in order to avoid the reductio whereby a thought about *any* x will result in

⁶ In Gennaro 1996, I call this the “wide intrinsicity view” (WIV). I also argue that Sartre held a similar view in Gennaro 2002. It is worth noting, however, that Brentano’s view is not quite the same as the WIV. In Gennaro (forthcoming) I further develop and defend the WIV against objections and show that it is importantly located between what I call Brentano’s “pure self-referentialism” and Rosenthal’s version of HOT theory.

x's being conscious. So the HOT theory does not really prove too much in this sense and various principled restrictions can be placed on the nature of both the lower-order and the meta-state in order to produce the mature theory. In this case, a rock is not a mental state and so having a thought about one will not render it conscious. After all, the HOT theory is attempting to explain *what makes a mental state* a conscious mental state. This is not properly recognized by those who put forward the problem of the rock.

There are also two further moves that can be made: First, it might be wise to raise the next natural question: What makes a state a *mental* state? There are differing views here, but one might, for example, insist that mental states must fill an appropriate causal-functional role in an organism (Armstrong 1981). Alternatively, one might even simply identify mental states with certain neural or bio-chemical processes in an organism (Crick 1994). Either way, however, it is clear that external objects, such as rocks, cannot meet these criteria. The lower-order states in question thus have certain special properties which make it the case that they become conscious when targeted by an appropriate HOT. It is also important to note that this response effectively handles other related objections to the HOT theory. Various *internal* states such as cancer (Dretske 1995, pp. 97, 100) and liver states (Block 1995, p. 280) are also ruled out by these criteria.⁷ Second, in a similar vein, if we return to the idea that the meta-state is an intrinsic part of a complex conscious state, then it is also clear that rocks cannot be rendered conscious by the appropriate HOT. This is because, on such a view, the meta-state must be more intimately connected with its object, and it is most natural to suppose that the target object must therefore be "in the head." That is, both "parts" of the complex conscious state must clearly be internal to the organism. Van Gulick (2000, 2004), who calls this "the generality problem," makes a similar point when he says that "having a thought...about a non-mental item such as the lamp on my desk does not make the lamp conscious...because [the lamp] cannot become a constituent of any such global [brain] state." (Van Gulick 2000, p. 301) Thus, it is difficult to compare the inner (mental)/inner (mental) relation as described by HO theories to the inner (mental)/outer (rock) relation described in Goldman's initial objection.

The problem remains, however, that this reply to the problem of the rock still invites a very natural response, as we have already seen above. Interestingly, such replies typically come in the form of "why-questions" or demands for further explanation: *Why* or *how* can such meta-states confer consciousness on their mental objects? Recall also part of the quote from Stubenberg above: "The worry is fueled by one's inability to comprehend *how* entering into this

⁷ For a brief but more recent reply from Rosenthal on Block's 'liver' version of the problem of the rock, see Rosenthal 2000a, p. 241.

relation is supposed to promote an unconscious state to consciousness.” (1998, p. 195, emphasis added) However, it is interesting to note that we now have a somewhat different, though importantly related, objection. Instead of claiming that HO theories prove too much, a question is now raised about how or if they can explain anything at all regarding consciousness. That is, the question has shifted from arguing that HO theories (if true) are too strong in the sense of making too many things conscious to objecting that they are much too weak to explain phenomenal consciousness. It is, in some ways, a curious counter-reply given the initial problem of the rock. This shift to a demand for a satisfying explanation will sound very familiar to those aware of what David Chalmers (1995) calls the “hard problem” of consciousness. It is now time to turn our attention to that problem, but one key point here is that the problem of the rock and the hard problem are really just two sides of the same coin.

III. The Hard Problem

A. The Challenge

David Chalmers (1995) presents the “hard problem” of consciousness⁸ as follows:

The really hard problem is the problem of *experience*....a subjective aspect....It is undeniable that some organisms are subjects of experience. But the question of *how* it is that [organisms] are subjects of experience is perplexing. *Why* is it that when our cognitive systems engage in visual and auditory information-processing, we have a visual or auditory experience...?...*How* can we explain *why* there is something it is like to entertain a mental image, or to experience an emotion? It is widely agreed that experience arises from a physical basis, but we have no good explanation of *why* and *how* it so arises. (1995, p. 201, emphases added after the first one)

The similarity between parts of this quote and any reasonable counter-reply to the problem of the rock should be obvious. The number of “why” and “how” questions in Chalmers’ quote echoes the demand for explanation that we saw at the end of the last section. Let me be clear about the connection between the initial problem of the rock and the hard problem. One way to put the dialectic thus far is as follows: Problem of the rock: “If HO theory

⁸ See also Shear 1997 where Chalmers acknowledges that this problem has been recognized by many philosophers in the past under different names. More recently, Seager (1999) calls his version of the hard problem “the generation problem.”

is true, why don't some objects of thoughts (e.g. rocks) become conscious when thoughts are directed at them?" HO theorist: "Well, rocks aren't mental states which are in the head and thus have importantly different properties than rocks, etc." Counter-reply (Hard Problem): "But exactly how or why does having HOTs directed at mental states make them conscious or explain conscious experience?" Now, to be sure, a version of this last question can even be found in Goldman's original objection (and is even clearer in the Stubenberg quotation), but this is precisely why I am arguing that the two objections are importantly connected and two sides of the same coin. Both objections rest on a challenge about what is needed for a theory of consciousness to render consciousness intelligible. Nonetheless, there is the subtle shift from addressing the nature of the *objects* of thoughts to the question of how or if invoking HOTs can provide a satisfying explanation of consciousness. Moreover, there is the move from objecting that the HOT theory, if true, is too strong, at least in the sense that it would make too many things conscious (e.g. rocks), to the complaint that the HOT theory, if true, does not present a strong enough explanation of phenomenal consciousness.

Now, in his 1995 paper, Chalmers is of course not mainly concerned with HO theories of consciousness; indeed, HO theories are noticeably absent from his treatment. Chalmers is instead concerned to show how various other attempts to explain consciousness have failed or merely address the relatively "easy problems" of consciousness, such as explaining the integration of information by a cognitive system or the ability of a cognitive system to access its own internal states. Thus, the bottom line objection is, once again, that various theories of consciousness do not explain enough.

It must first be acknowledged that any HO theorist should be willing to take up this challenge. We must face the hard problem head on, though we may never fully satisfy those who advance it. Any theory of consciousness should at least attempt to explain the "what it is like" (Nagel 1974) and qualitative character of phenomenal experience. This has perhaps not always been the case. Stubenberg (1998), for example, alleges that HO theorists have intentionally avoided this type of problem or considered it to be unimportant. He argues that HO theories do not explain or even address *the nature* of qualitative experience. Stubenberg uses HO theorist's own words against them. For example, Lycan says that "I am not here addressing issues of qualia or phenomenal character...There may be Inner Sense theorists who believe that their views solve problems of qualia; I make no such claim, for I think qualia problems and the nature of conscious awareness are mutually independent..." (1996, p. 15) Stubenberg also cites Rosenthal's view that "the properties of being conscious and having sensory quality are independent of one another, and a satisfactory account of each property requires us to investigate them separately." (1991, p. 16)

While Lycan and Rosenthal may seem to contradict such remarks elsewhere,⁹ I agree with Stubenberg that many HO theorists have been slow to address the hard problem and have often avoided it intentionally.¹⁰ However, I also disagree with him in one important way: As I will argue below, the HOT theory can explain qualitative experience even if some HOT theorists, such as Rosenthal, argue for the independence of consciousness and what he calls ‘sensory qualities.’ Unlike myself, Rosenthal holds, for example, that there are unconscious sensory qualities. But it surely does not follow that, when the sensory qualities *are* conscious, they cannot be explained in terms of the HOT theory.

B. A Proposed Solution

Before offering an answer to the hard problem, it is crucial to distinguish between various theories of consciousness. Some theories are entirely nonreductionist; for example, they treat conscious experience as a fundamental or an irreducible feature of the world (see e.g. Chalmers 1995, 1996) However, the HOT theory is reductionist in the sense that consciousness is to be

⁹ For example, Stubenberg cites Lycan’s surprising statement that the “inner-sense account affords the best known solution I know to the problem of subjectivity and ‘knowing what it’s like’...” (Lycan 1996, p. 15). See also Rosenthal 2002, section IV. However, to be fair to both Lycan and Rosenthal, it is absolutely crucial to separate their explanations of “sensory” or “qualitative” properties *as they use these terms*, which are offered in a way so as to allow for these properties to be unconscious (see additionally e.g. Lycan 1996, pp. 75-77, and Rosenthal 2000a, pp. 235-6), from their explanations of “knowing what it is like” to experience those properties from the first-person point of view. It is primarily with the latter that their respective versions of HO theory are invoked. For my own part, I choose to reserve the terms ‘sensory’ and ‘qualitative’ for the subjective first person aspects of consciousness, but this is largely (though not entirely) a terminological difference. However, part of the point of the rest of this paper will be to show that the HOT theory can be used to explain consciousness and to address the hard problem in ways that go beyond what other HO theorists have said up to this point.

¹⁰ Though, as I will urge below, I would at least like to think that I did implicitly address the hard problem in Gennaro 1996, especially chapters 3 and 4.

explained in terms of something else, i.e. nonconscious mental states. It is important to note, though, that the HOT theory does not attempt to explain consciousness in *nonmentalistic* or “naturalist” terms. Of course, HOT theorists, including myself, tend to be token-identity materialists in the end, but we prefer to leave that empirical question for a separate second step reduction to be filled in later by brain science. I believe that the HOT theory provides necessary and sufficient conditions for what makes a mental state conscious, but whatever realizes that theory *in our brains* is a separate empirical question. On the other hand, other theories of consciousness are reductionist in the stronger sense that they attempt to explain consciousness *directly* in physical or neurophysiological terms (e.g. Crick and Koch 1990, Crick 1994). Chalmers is of course well aware of these distinctions, but he argues that all such attempts still fail to address the hard problem. The quote above from Chalmers does show, however, how statements of the hard problem can sometimes gloss over these differences. Nonetheless, the key common general question is: How exactly does x (where ‘x’ is invoked by some alleged theory of consciousness) explain how consciousness arises from the presence of x? Recall that we even encountered a somewhat similar question in connection with the problem of the rock.

Now, according to the HOT theory, the ‘x’ is to be filled in with ‘the presence of a suitable higher-order thought.’ But then, of course, the hard problem asks: *Why* or *how* exactly does the presence of such a HOT result in a conscious mental state? Some of what I have said in a previous publication (Gennaro 1996) was an indirect attempt to answer this question. However, at that time, I did not explicitly have the so-called “hard problem” in mind. I want to show here how that basic idea can be brought to bear on this problem.

The solution, then, is that HOTs explain how consciousness arises because the *concepts* that figure into the HOTs are presupposed in conscious experience. Let us stick to first-order perceptual states. In very much a Kantian spirit, the idea is that first we passively receive information via our senses. This occurs in what Kant (1781/1965) calls our “faculty of sensibility.” Some of this information will then rise to the level of unconscious mental states which can, of course, also cause our behavior in various ways. But such mental states do not become conscious until the “faculty of understanding” operates on them via the application of concepts. I contend that we should understand such concept application in terms of HOTs directed at the incoming information. Thus, I consciously experience the brown tree *as a brown tree* partly because I apply the concepts ‘brown’ and ‘tree’ (in my HOTs) to the incoming information via my visual perceptual apparatus. More specifically, I have a HOT such as “I am seeing a brown tree now.” Kant, of course, similarly urges that it takes the cooperation of both the sensibility and understanding to produce conscious experience. It is crucial to remember that these HOTs are not themselves conscious; thus, HOTs and their concepts are, we might say, “presupposed” in conscious experience. The understanding unconsciously “synthesizes” the raw

data of experience. To be sure, even nonconscious mental states also involve some form of conceptualization or categorization in so far as they have intentional content. However, part of the motivation for the HOT theory is to explain when and how a nonconscious state becomes conscious, and the answer on the HOT theory is that the subject becomes aware of the state, i.e. has a HOT directed at it. A HOT theorist will therefore argue that, for example, a purely first-order (FO) account of consciousness (e.g. Dretske 1995, Tye 1995) cannot adequately or equally explain the difference between nonconscious and conscious mental states. Hence, the concepts in question must be in the HOTs and they are responsible for the “what it is like” nature of qualitative experience. A full explanation of all the advantages of HO theory is beyond the scope of this paper, but see, for example, Carruthers (2000, chapters five and six) for an excellent discussion of why one should opt for a HO theory over a FO account. Moreover, the HO theorist will argue that reductionism in mentalistic terms is preferable to reductionism in nonmental terms.¹¹

Rosenthal has also more recently argued in a somewhat similar Kantian fashion. Beginning with his well-known example of wine-tasting, he first says:

Learning new concepts for our experiences of the gustatory and olfactory properties of wines typically leads to our being conscious of more fine-grained differences among the qualities of our sensory states. Similarly with other sensory modalities...new concepts appear to generate new conscious sensory qualities. (2002, p. 413)

But Rosenthal then uses the example of hearing the sound of an oboe. He argues that if we systematically removed *all* of the relevant concepts (e.g. ‘sound of a woodwind’) involved in having that conscious experience, then there would no longer be the conscious experience. This seems true for any conscious experience. As Rosenthal puts it, it is “plausible that peeling away that weakest HOT would result, finally, in its no longer being like anything at all to have that sensation.” (2002, p. 414) The concepts that we have clearly color the very experiences we have and removing all of them would eliminate the experience itself. Indeed, having such concepts, on the HOT theory is both necessary and sufficient for having subjective conscious experience.

Now I submit that this is an adequate answer to the above challenge, aside perhaps from a bit more detail.¹² Of course, some will no doubt always remain dissatisfied and will still want to

¹¹ See Gennaro 1996, pp. 12-15, and section IVB below.

¹² For more detail, see Gennaro 1996 chapters 3 and 4.

ask a further question: *Why* does the higher-order application of concepts give rise to conscious experience? But this, I suggest, is not a legitimate question. We have already reached the rock bottom brute fact about the way that conscious minds work, and the chain of explanation has already come to an end. The Kantian idea that concepts make our experience of the world possible is surely a widely held view about the nature of conscious experience. I do not think that it makes sense to ask why this is so. As Strawson puts it: “[if] any item is even to enter our conscious experience we must be able to classify it in some way, to recognize it as possessing some general characteristics.” (1966, p. 20)¹³ Notice that this solution is unlike reductionist accounts in *nonmentalistic* terms, and so it is immune from Chalmers’ criticism about the plausibility of those theories. For example, there is no problem about how some specific *brain activity* produces conscious experience. Chalmers’ criticism that *functional* explanations are inadequate because one can always ask “*Why is the performance of these functions accompanied by experience?*” (1995, p. 203) is equally beside the point. The HOT theory is not a functional explanation which merely addresses the easy problems of consciousness. In any case, the HOT theory contends that a reductionist theory of consciousness can be provided in mentalistic terms in a way that can solve the hard problem. This is also, then, the answer to those who demand a further explanation as a counter-reply to the problem of the rock.

There are, of course, other questions regarding concepts that do make sense: What are concepts? What is it to possess a concept? How do we apply concepts? How do we acquire concepts? Which, if any, concepts are innate? How are concepts realized in the brain? These are notoriously difficult questions to answer, each with a long history of failed or questionable attempts. Concepts are perhaps best understood as universals which fix the extension of terms,

¹³ No doubt some will here be tempted to respond with a belief in some version of the so-called “non-conceptual content” in experience, which has recently become quite a hot topic. Needless to say, I am firmly in the conceptualist camp. However, a full response to this line of argument will have to wait for another occasion, but see Gunther (2003) for an excellent anthology on the topic. My main purpose here is to bring this Kantian-style version of the HOT theory to bear on the hard problem. Recall also that a HOT theorist argues that purely first-order (FO) accounts of consciousness (e.g. Dretske 1995, Tye 1995) cannot adequately or equally explain the difference between nonconscious and conscious mental states (see, for example, Carruthers 2000, chapters five and six). This is again why the concepts in question must be in the HOTs themselves.

but answering the remaining questions is even more difficult. For example, the search for necessary and sufficient conditions of *application* for any given concept often seems doomed to failure. Analyzing concept possession in terms of, say, mental images or mere behavioral discrimination seems insufficient for various well-known reasons. This is not the place to put forth a theory of concepts,¹⁴ but my point here concerning the HOT theory is simply that these legitimate questions are far different in structure than the earlier illegitimate question raised in relation to the hard problem, i.e. *Why* does the higher-order application of concepts give rise to conscious experience? In other words, it may be very difficult to explain how we acquire or apply concepts, but that is different than demanding a further explanation for the fact that, *given the presence of concepts (in the HOTs)*, conscious experience results. If there is a real hard problem of consciousness, I suggest that it has more to do with concept acquisition and application.

As I mentioned earlier, the HOT theory is notoriously absent from Chalmers' 1995 discussion, and it is only briefly mentioned in his 1996 book. Perhaps he would place the HOT theory in the same category as others that approach the problem of consciousness at the level of cognitive psychology. For example, Baars' (1988) global workspace theory of consciousness and Dennett's (1991) multiple drafts model could be viewed as similar to the HOT theory in some respects. But Chalmers urges us to agree that neither of these theories really addresses the hard problem. Baars leaves unanswered the question: *Why* is the information in the global workspace experienced? And Dennett's theory "is largely directed at explaining the reportability of certain mental contents." (1995, p. 205) Thus, such theories never really do address the hard problem or really only address an "easy" problem. Even if Chalmers is right in his criticism of Baars and Dennett, I have tried to show how the HOT theory is in a better position to address the hard problem. The HOT theory is not merely dealing with an easy problem (e.g. in terms of some purely functional relation); nor is it ignoring the hard problem altogether. It is attempting to explain the very *structure* of conscious mental states as well as *how* such states come to have their qualitative feel.

Interestingly, Chalmers makes several remarks akin to the HOT theory in presenting what he calls "naturalistic dualism," which is the view that experience is a basic or fundamental feature of the universe over and above the properties invoked by physics. For example, when presenting his "principle of structural coherence," Chalmers (1995) describes a close connection between "consciousness" and "awareness." (cf. Chalmers 1996, pp. 218-29) In his response to

¹⁴ But see Gennaro 1996, pp. 54-68 for an attempt to address some of these questions. Also, see Margolis and Laurence 1999 for a sampling of research on the nature of concepts.

various authors (in Shear 1997), Chalmers also says that “I find it plausible that there is an intimate relationship between consciousness and thought.” (p. 396) It is clear that this is a view shared by HOT theorists. A higher-order thought (or awareness) of a lower-order state renders that state conscious. I am not suggesting that Chalmers is a closet HOT theorist, but such quotes reveal an important agreement between him and the HOT theory. Instead, Chalmers ultimately uses the notion of “information” in defending a “double-aspect” theory whereby information “has two basic aspects, a physical aspect and a phenomenal aspect.” (1995, p. 216)

C. Conceivability, Necessity, and the HOT theory

Now, I do not want to give the impression that I agree with well-known “conceivability” and “zombie” arguments against physicalism. The idea here is basically that if it is conceptually possible for there to be a physically identical creature to me that lacks conscious experience, then ultimately materialism is false. My point here will be only that the HOT theory is in better shape to deal with such arguments and can even avoid them entirely. As a matter of fact, I am quite sympathetic to those who have responded in a variety of ways to Chalmers’ use of such arguments (see e.g. Papineau 1998, Block and Stalnaker 1999, Loar 1999, Yablo 1999, Carruthers 2000, Perry 2001, Botterell 2001). In particular, it is not clear to me how any ontological conclusion about the actual world can be drawn from such arguments. I have little to add here to those standard criticisms of the conceivability argument, and, in any case, that is a major topic which goes beyond the scope of this paper. But even Joseph Levine, a key proponent of the view that there is an “explanatory gap” between conscious experience and the physical world, admits that the “explanatory gap is primarily an epistemological problem, not necessarily a metaphysical one.” (2001, p. 10)

Still, Chalmers might object to my solution by asking “How can the HOT theory express a necessary truth, especially since its denial does not seem to involve a contradiction?” “Even if consciousness necessarily presupposes concepts, why must those concepts be constituents of *higher-order* thoughts?”¹⁵ Surely, he might say, it is *possible* for the HOT theory to be false, and, if so, then the rest of his argument can proceed from there. This line of thought is clear from the following key passage about the possibility of zombies:

¹⁵ Once again, explanations of the advantages of HO theory over FO theory have been given elsewhere (e.g. Carruthers 2000); and see also section IIIB and note 13. My point here is not to argue for the HOT theory anew; instead, it is only to address the Chalmers-style challenge that the denial of the HOT theory does not seem to result in an explicit contradiction.

the conceivability of zombies seems...obvious to me...While this possibility is probably empirically impossible, it certainly seems that a coherent situation is described; I can discern *no contradiction* in the description. In some ways an assertion of this logical possibility comes down to brute intuition but...[a]lmost everybody, it seems to me, is capable of conceiving this possibility. (1996, p. 96, emphasis added)

It seems to me that there are at least the following three responses open to the HOT theorist: *First*, to the extent that Chalmers is willing to allow “brute intuition” to guide possibility, then I must admit that I do indeed find it very difficult to conceive of a creature with conscious experience but without HOTs, or vice versa. This is based on my critical and substantial reflection on the nature of conscious mental states, which should count for something. More importantly, however, even if I do admit the possibility of consciousness without HOTs (or vice versa), it would seem that something similar is also true of Chalmers’ own view. To the extent that Chalmers is merely engaged in “conceptual analysis,” surely it is not *contradictory* to suppose that, for example, consciousness is not a fundamental ontological feature of the world. Surely it is also not *contradictory* to suppose that consciousness is not tied to “information” in the way that Chalmers speculates. Surely, then, there are possible worlds where Chalmers’ view is false. Perhaps he would deny such a possibility, but then I think I can do the same with respect to the HOT theory. On the other hand, if Chalmers accepts this possibility, then presumably he would not take it as proving that his ontological view about the actual world is false. And so I should again be permitted to say the same about the HOT theory; after all, if that is good enough for Chalmers, then it is good enough for me. It may be that Chalmers is more concerned with the *ontology* of consciousness in this context instead of mere conceptual analysis, and so he might object to treating his view as mere conceptual analysis. But then, once again, the same could go for proponents of the HOT theory as well as for various naturalist theories. In all of these views, including Chalmers’ own, we should then wonder how conceiving of their falsity can force us to conclude that they are false as an ontological view about the actual world.

A *second* more provocative reply is much more involved. It requires us first to define an often used, but sometimes conflated, threefold distinction. There are analytic and synthetic *statements* or *sentences*; the former we can define as a sentence whose denial is contradictory whereas the latter are simply sentences which are not analytic. These terms are usually understood as making a claim about the *meaning* of various concepts involved in sentences, as in the classic analytic example of “all bachelors are unmarried.” There is also *a priori* and *a posteriori* (or empirical) *knowledge*: the former can be known independently of sense experience

whereas the latter is acquired via the senses. Finally, there are contingent and necessary *truths*; the former are true in this and (perhaps) some other possible worlds, and the latter are true in all possible worlds.

The problem with the objection, then, is that it assumes that “only analytic statements can express necessary truths.” Now, first of all, the familiar “water is H₂O” example shows that this assumption is false (Kripke 1972). Of course, Chalmers and others know this very well but they then typically argue at great length about the alleged differences between the water-H₂O case and consciousness. Indeed, this is the main purpose behind Chalmers’ so-called “two-dimensional semantics” where he distinguishes between the primary and secondary intensions (meanings) of a concept. However, what really causes trouble for his opponents here is the fact that “water is H₂O” is known a posteriori.¹⁶ After all, Chalmers makes it clear that his main targets are those proposed “solutions” to the hard problem which mention neurophysiological properties. This is what Botterell (2001, pp. 22-3) aptly calls “a posteriori physicalism” and then quotes Chalmers (from 1999, p. 474) as having in mind the view that “[psychological nature is] not necessitated a priori by physical [nature], but...that [it is] necessitated a posteriori by physical [nature].” Botterell rightly characterizes this view as follows: “[i]n short, a posteriori physicalism maintains that although there is a *necessary* entailment of psychological nature by physical nature, there is no *a priori* or *conceptual* connection between the two.” (2001, p. 23)

My point here, however, is that whatever value Chalmers’ views have in relation to a posteriori physicalism, they do not apply to the HOT theory. Unlike a posteriori physicalism, the HOT theory does hold that the relationship between the explanandum and explanans is a priori and, in this sense, a conceptual one. There is a *logical entailment* between the content of the

¹⁶ Witness Chalmers remark that he is outlining “a two-dimensional intensional framework for handling a posteriori necessity.” (1999, p. 435) For those unfamiliar with Chalmers’ terminology here and for more details on the primary/secondary intensions distinction, see his 1996 pp. 52-71. For example, Chalmers explains that “the primary intension picks out a referent of a concept in a world when it is *considered as actual*...whereas the secondary intension picks out the referent of a concept in a world when it is *considered as counterfactual*, given that the actual world of the thinker is already fixed.” (1996, p. 60) Perhaps most relevant here, however, is when Chalmers says that “[w]e might as well think of the primary and secondary intensions as the a priori and a posteriori aspects of meaning, respectively.” (1996, p. 62)

HOT theory and statements about the nature of conscious experience, and Chalmers agrees that “if B-properties are logically supervenient on A-properties according to primary intensions, then the implication from A-facts to B-facts will be a priori.” (1996, p.70) This interpretation of the HOT theory is also supported by Van Gulick when he says that it proposes to offer “logically sufficient condition[s] from which one could deduce the existence and nature of the relevant feature of consciousness.” (Van Gulick 1995, p. 70) Some of this a priori reasoning can be found above, but is also presented in more detail by any number of HO theorists.¹⁷

Thus, even if I am willing to concede that *the statement* of the HOT theory is synthetic simply because its denial does not seem to result in an explicit contradiction, we still can reply as follows: If there has ever been any dispute about whether or not synthetic a posteriori truths can be necessary, it is normally due to the fact that they are known a posteriori. But if at least the fundamental thesis of the HOT theory can be known a priori (as I suggest), then, even according

¹⁷ See, for example, Rosenthal 1986, 1990; Gennaro 1996; and Lycan 2001. It might be objected that the HO hypothesis is generally put forth as an empirical (as opposed to an *a priori*) claim. This is certainly true. However, I think there is a real issue here for a number of reasons. First, if ‘an empirical claim’ means “in principle empirically falsifiable” or “consistent with and sometimes supported by empirical and scientific evidence,” then I certainly would agree. But, second, the idea that the basic truth of the HOT theory is known (or at least claimed to be known) *a priori* seems consistent with it as “empirical” in those senses. We might claim to know that p is true *a priori*, but then, for example, look to see if any empirical evidence can support or falsify it. Of course, if it is falsified, then we would say that we never really knew it in the first place. Third, it seems to me that at least some of the claims to know HO theory are, indeed, based on *a priori* reasoning, contrary to what HO theorists might otherwise say. Lycan’s (2001) argument comes to mind as does Rosenthal’s (and my) frequent assertion about the “intuitively” obvious truth that conscious mental states are those states that we are aware that we are in. Indeed, this truth is treated as a fundamental “definition” in Lycan 2001, and Bill Lycan himself has told me (in correspondence) that he wonders if HO theories are “nearly trivially true.” In any case, if I differ from other HO theorists on the extent to which HO theory is known *a priori*, then so be it.

to Chalmers, the idea of an a priori necessary truth would be far less controversial. The HOT theory can, therefore, embody a synthetic, a priori, and necessary truth. If one agrees with Kant (as I do) that there are synthetic a priori truths (such as mathematical truths, ‘every event has a cause’), then the HOT theory can again avoid Chalmers’ arguments against a posteriori physicalism. Logical entailments can be found in the HOT theory in a way that is absent in purely physicalist theories of consciousness. Thus, even if the *statement* of the HOT theory is synthetic, we can still maintain that it is *known* a priori and that it *expresses* a necessary truth. At the very least, then, Chalmers is unjustly ruling out synthetic a priori truths, though he rightly recognizes the close link between those truths known a priori and their metaphysical necessity.

It may seem odd to talk about the HOT theory as a synthetic statement together with the notion that there are logical entailments between it and statements about conscious experience. But there is really no problem here. Suppose Kant was right about the synthetic nature of statements in geometry. It would still make perfect sense to talk about logical entailments between such statements and other statements in geometry or even in other disciplines. Analogously, even if the main claim of the HOT theory is synthetic, we can still make sense of logical entailments between a particular example of the HOT theory at work and the corresponding statement about the nature of that particular conscious experience.

Third, a HOT theorist might, of course, even agree with Quine (1951) and question the very distinction between analytic and synthetic statements. At the very least, it could even be argued that some apparently synthetic statements are “really” analytic in the sense that, to use Kant’s phrase, they contain predicates which are *implicitly* contained in the subject. That is, perhaps one can admit that the denial of a sentence does not lead to an *explicit* or *transparent* contradiction, but it nonetheless does eventually lead to an *implicit* one. This line is not always so easy to draw. After all, even such clear classic cases as “all bachelors are unmarried” involve some unpacking of the notion of ‘bachelor.’ Thus, a HOT theorist may even wish to maintain that, say, “being accompanied by a HOT” is implicitly contained in the *concept* “conscious mental state.” As Chalmers himself rightly points out: “[i]n certain cases, the decision about what a concept refers to in the actual world involves a large amount of reflection about what is the most reasonable thing to say...” (1996, p. 58) Chalmers then goes on to admit that such a process is still best understood as engaging in a priori reasoning (1996, pp. 58-9). In the end, however, what is most important to remember is that we are identifying the two *properties* in question. Analyzing the *concepts* is one thing, but it is the ontological identity between *properties* that is most central (Loar 1999).

IV. Two Applications

A. First Application: Van Gulick’s Approach

It will be helpful to place this entire discussion in the context of a paper by Robert Van Gulick (1995). He asks the meta-philosophical question “What would count as explaining consciousness?” and his purpose is to “untangle and clarify the various distinct issues which sometimes get run together...” (1995, p. 61) He accomplishes this by distinguishing (A) various *explananda* of consciousness (i.e. what needs to be explained); (B) the domain of *explanans* (i.e. in what terms we might construct an explanation); and (C) a *linking relation* that must hold between the explanandum and the explanans in order to provide a satisfactory explanation. Under (A) Van Gulick lists five features: (A1) the un/conscious state distinction, (A2) non/conscious creature distinction, (A3) qualitative aspect, (A4) phenomenal aspect, and (A5) subjectivity. Under (B) he lists the following: (B1) physical, (B2) functional, (B3) naturalistic, (B4) nonconscious mental states. And under (C) Van Gulick mentions (C1) logical sufficiency, (C2) nomic sufficiency, (C3) intuitive sufficiency, and (C4) predictive and pragmatically useful modelling.

Van Gulick points out how this menu generates ninety-six possible interpretations of his original ambiguous question. Furthermore, he argues that some theories are clearly only concerned with one or two features from each category. For example, a hardcore reductionist approach would presumably only use B1 or B3 and probably C2 (or perhaps C4) in order to explain one or more items under (A). Although Van Gulick does not demand that a theory of consciousness explain all of the explananda he lists under (A), it seems reasonable to hold that any good theory of consciousness should be able to do so. Presumably, this is precisely the force behind the hard problem. In particular, a theory of consciousness should be able to explain A3-A5 (qualitative, phenomenal, subjectivity). Although Van Gulick makes several subtle and somewhat terminological distinctions between those three explananda, he acknowledges that they could all be grouped under the general heading of “first person qualitative experience.”

Against this backdrop I wish to make several points regarding the HOT theory and the hard problem: First, it is interesting to note that (A) differs from (B) and (C) in one important way; namely, that all of A’s features should be explained by any good theory of consciousness whereas we are not required to pick more than one item from either of the other two categories. Indeed, for example, using B1 as the explanans (physical) would even rule out using B4 (nonconscious mental states).

Second, Van Gulick rightly notes that the HOT theory is probably most often understood as explaining A1; that is, the un/conscious state distinction. The HOT theory can also presumably handle A2 (non/conscious creature distinction) fairly well. However, as I have argued, we can and should also explicitly take up the challenge of explaining A3-A5, and, after all, part of the point of this paper is to show how the HOT theory can explain qualitative consciousness.

Third, in the domain of explanans, the HOT theory clearly relies on a B4 approach; that is, explain consciousness in terms of nonconscious mental states. Again, there is no requirement for a theory of consciousness to use more than one explanans from the B list in order to explain consciousness. Indeed, it is crucial to note that Van Gulick separates B4 from both functional and physical explanations (even though they all could be used together). As I argued earlier, then, the HOT theory is immune from the criticisms leveled against these two approaches by Chalmers.

Fourth, Van Gulick associates the HOT theory with C1; that is, the linking relation of logical sufficiency. I agree with this as should be clear from my earlier remarks. Once again, recall that Van Gulick rightly states that the HOT theory does propose to offer “logically sufficient condition[s] from which one could deduce the existence and nature of the relevant feature of consciousness.” (Van Gulick 1995, p. 70) Indeed, I have argued that the HOT theory provides necessary and sufficient conditions for a mental state to be conscious. The HOT theory expresses a necessary truth about what makes a mental state conscious (which we can come to know *a priori*). There *could not* be a conscious mental state which is presented to its owner devoid of *all* conceptualization. A conscious state *must* be presented to its owner in some way or other; that is, it must be thought of under some mode of presentation. So again, then, if the HOT theory is defensible, it is immune from Chalmers’ arguments against physicalist approaches, particularly when he argues, via so-called “zombie arguments,” that such a physicalist explanans does not *logically entail* the presence of conscious experience.

Fifth, the HOT theorist can also appeal to the linking relation that Van Gulick calls “intuitive sufficiency” (C3). There is, I believe, something uniquely simple and intuitive about the HOT theory’s ability to explain consciousness. Of course, as Van Gulick recognizes, “the problem [here]...is what strikes us as intuitive is highly context sensitive and relative.” (1995, p. 71) I am normally not one to rely heavily on intuition. Nonetheless, I do think that the HOT theory has an intuitive appeal over other accounts for the very reasons given earlier: it can accommodate the intuitive ideas that we are aware of the conscious states we are in and that concepts are presupposed in conscious experience. Moreover, the HOT theory preserves these intuitions while avoiding the problems with purely physicalist accounts which, Chalmers argues, cannot properly respond to the “intuition” that for any brain process “it is conceptually coherent that it could be instantiated in the absence of experience.” (1995, p. 208)

B. Second Application: Schröder’s Argument

Jürgen Schröder (2001) argues that reductionist accounts of consciousness in physical or “naturalist” terms have greater explanatory power than HOT theories. He addresses a number of areas that I cannot delve into here, such as the unity of consciousness, the inseparability of qualia

and consciousness, and the differentiation of qualia. However, if my argument thus far has been successful, there is one key area that has been ignored by Schröder; namely, that HO theories can provide a better explanation of the logically sufficient conditions of consciousness. That is, unlike naturalist accounts, the HOT theory can avoid the kinds of concerns, raised by Chalmers and others, about the logical relationship between explanandum and explanans. I suggest that this is a key advantage over naturalist accounts.

Now Schröder does address one related issue. He responds to the potential criticism that HO theories can provide *necessary* conditions whereas naturalist accounts cannot because the former allow for the multiple realizability of conscious states in a way that the latter cannot. For example, if a naturalist theory is couched in neurophysiological terms, then it apparently cannot offer a necessary condition for consciousness because it seems possible for other organisms to have conscious states and yet not have the corresponding neurophysiology. Schröder responds, in part, by suggesting that, say, if a non-HO theory “identifies consciousness with the stability of activation vectors, this property – stability of activation – can be regarded as a relatively high-level property because nothing is determined about the kinds of units that are activated or about the specific form of energy that a state of activation is based on.” (2001, p. 29) Although I am doubtful that his reply is fully adequate in the end, I do not wish to debate Schröder on this point here. I raise it only to point out an important contrast.

Let us even grant that Schröder can answer that criticism. Unfortunately, he would still not be addressing the key advantage of HO theories raised in this paper; namely, *logical sufficiency from explanans to explanandum*. Schröder’s response to the above criticism about multiple realizability deals with a *necessary* condition of the explanandum. The “direction” of explanation here, however, is quite different. We might say that Schröder is responding to a concern about the relation *from explanandum to explanans* whereas, in this paper, we have been more concerned with the (logically sufficient) relation *from explanans to explanandum*. I do not mean to suggest that Schröder intended to address this issue and that he has failed to do so successfully. He simply never raises it at all. My point, then, is that if I am right, he has altogether ignored one important further advantage of HO theories over naturalist accounts. The HOT theory, especially when viewed as a *philosophical* theory, is preferable to naturalist accounts of consciousness.¹⁸

V. Conclusion

¹⁸ And on the HOT theory’s ability to explain at least our *sense* of the unity of consciousness, see Rosenthal 2000b and 2003.

I conclude that the HOT theory can be successfully defended against both the problem of the rock and the charge that it does not address the hard problem of consciousness. The HOT theory is therefore not really between a rock and a hard place. The HOT theory does not claim that a meta-state about *anything* will render it conscious: the target of the HOT must be a suitable mental state. Furthermore, analyzing the HOT theory in terms of concept application reveals that it can address the hard problem. One advantage of this approach is that the HOT theory is then immune to the criticisms that Chalmers levels against a posteriori physicalism. It turns out, though, that the two objections raised are really two sides of the same coin. The charge of not answering the hard problem is really just a typical counter-reply to the problem of the rock. Finally, we saw how the approach advanced here can be usefully applied to two important recent articles dealing with the general problem of explaining consciousness and the explanatory power of the HOT theory in particular.¹⁹

References:

- Armstrong, D. (1981), *The Nature of Mind and Other Essays* (Ithaca, NY: Cornell).
- Baars, B. (1988), *A Cognitive Theory of Consciousness* (Cambridge: Cambridge University Press).
- Block, N. (1995), 'On a confusion about a function of consciousness', *Behavioral and Brain Sciences*, 18, pp. 227-87.
- Block, N. and Stalnaker, R. (1999), 'Conceptual analysis, dualism, and the explanatory gap', *The Philosophical Review*, 108, pp. 1-46.
- Botterell, A. (2001), 'Conceiving what is not there', *Journal of Consciousness Studies*, 8 (8), pp. 21-42.
- Brentano, F. (1874/1973), *Psychology From an Empirical Standpoint* (New York: Humanities).
- Byrne, A. (1997), 'Some like it HOT: Consciousness and higher-order thoughts', *Philosophical Studies*, 86, pp. 103-29.
- Carruthers, P. (2000), *Phenomenal Consciousness* (New York: Cambridge University Press).

¹⁹ I would like to thank Bill Lycan, Leopold Stubenberg, Robert Lurz, Ken Williford, Peter Carruthers, Frederick Adams, and two anonymous reviewers for the *Journal of Consciousness Studies* for many helpful comments on an earlier version of this paper. A shorter version of this paper was delivered at the 2003 APA Central Division Meetings in Cleveland, Ohio; and at the 2002 "Toward a Science of Consciousness" Conference in Tucson, Arizona.

- Chalmers, D.J. (1995), 'Facing up to the problem of consciousness', *Journal of Consciousness Studies*, 2, pp. 200-19.
- Chalmers, D.J. (1996), *The Conscious Mind* (New York: Oxford University Press).
- Chalmers, D.J. (1999), 'Materialism and the metaphysics of modality', *Philosophy and Phenomenological Research*, 59, pp. 473-96.
- Crick, F. (1994), *The Astonishing Hypothesis: The Scientific Search for the Soul* (New York: Scribner).
- Crick, F. and Koch, C. (1990), 'Towards a neurobiological theory of consciousness', *Seminars in the Neurosciences*, 2, pp. 263-75.
- Dennett, D.C. (1991), *Consciousness Explained* (Boston: Little Brown).
- Dretske, F. (1995), *Naturalizing the mind* (Cambridge, MA: MIT Press).
- Gennaro, R. (1996), *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness* (Amsterdam: John Benjamins).
- Gennaro, R. (2002), 'Jean-Paul Sartre and the HOT theory of consciousness', *Canadian Journal of Philosophy*, 32, pp. 293-330.
- Gennaro, R. (ed. 2004), *Higher-Order Theories of Consciousness* (Amsterdam: John Benjamins).
- Gennaro, R. (forthcoming), 'Between pure self-referentialism and the (extrinsic) HOT theory of consciousness', in *Consciousness and Self-Reference*, eds. Uriah Kriegel and Kenneth Williford (Cambridge, MA: MIT Press).
- Goldman, A. (1993), 'Consciousness, folk psychology, and cognitive science', *Consciousness and Cognition*, 2, pp. 364-82.
- Gunther, Y. (ed. 2003), *Essays on Nonconceptual Content* (Cambridge, MA: MIT Press).
- Kant, I. (1781/1965), *Critique of Pure Reason*, trans. Norman Kemp Smith (New York: St. Martin's Press).
- Kripke, S. (1972), *Naming and Necessity* (Cambridge, MA: Harvard).
- Levine, J. (2001), *Purple Haze: The Puzzle of Consciousness* (New York and Oxford: Oxford University Press).
- Loar, B. (1999), 'David Chalmers's *The Conscious Mind*', *Philosophy and Phenomenological Research*, 59, pp. 465-72.
- Lycan, W.G. (1996), *Consciousness and Experience* (Cambridge, MA: MIT Press).
- Lycan, W.G. (2001), 'A simple argument for a higher-order representation theory of consciousness', *Analysis*, 61, pp. 3-4.
- Margolis, E. and Laurence, S. (eds. 1999), *Concepts: Core Readings* (Cambridge, MA: MIT Press).
- Nagel, T. (1974), 'What is it like to be a bat?', *Philosophical Review*, 93, pp. 435-50.

- Papineau, D. (1998), 'Mind the gap', in *Language, Mind, and Ontology. Philosophical Perspectives vol. 12*, ed. J. Tomberlin (Oxford: Blackwell), pp. 373-88.
- Perry, J. (2001), *Knowledge, Possibility, and Consciousness* (Cambridge, MA: MIT Press).
- Quine, W.V. (1951), 'Two dogmas of empiricism', *Philosophical Review*, 60, pp. 20-43.
- Rosenthal, D. (1986), 'Two concepts of consciousness', *Philosophical Studies*, 49, pp. 329-59.
- Rosenthal, D. (1990), 'A theory of consciousness', Report No. 40 on MIND and BRAIN, Perspectives in Theoretical Psychology and the Philosophy of Mind (ZiF), University of Bielefeld. A version of this paper is reprinted in *The Nature of Consciousness: Philosophical Debates*, (eds. 1997) N. Block, O. Flanagan, and G. Güzeldere (Cambridge, MA: MIT Press).
- Rosenthal, D. (1991), 'The independence of consciousness and sensory quality', in *Consciousness*, ed. E. Villanueva (Atascadero, CA: Ridgeview), pp. 15-36.
- Rosenthal, D. (1993), 'State consciousness and transitive consciousness', *Consciousness and Cognition*, 2, pp. 355-63.
- Rosenthal, D. (2000a), 'Metacognition and higher-order thoughts', *Consciousness and Cognition*, 9, pp. 231-42.
- Rosenthal, D. (2000b), 'Introspection and self-interpretation', *Philosophical Topics*, 28, pp. 201-33.
- Rosenthal, D. (2002), 'Explaining consciousness', in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers (New York: Oxford University Press), pp. 406-21.
- Rosenthal, D. (2003), 'Unity of consciousness and the self', *Proceedings of the Aristotelian Society*, 103, pp. 325-52.
- Rosenthal, D. (2004), 'Varieties of higher-order theory', in Gennaro 2004.
- Schröder, J. (2001), 'Higher order thought and naturalist accounts of consciousness', *Journal of Consciousness Studies*, 8 (11), pp. 27-46.
- Shear, J. (ed. 1997), *Explaining Consciousness* (Cambridge, MA: MIT Press).
- Seager, W. (1999), *Theories of Consciousness* (New York and London: Routledge).
- Stubenberg, L. (1998), *Consciousness and Qualia* (Amsterdam: John Benjamins).
- Strawson, P.F. (1966), *The Bounds of Sense* (New York and London: Methuen).
- Tye, M. (1995), *Ten Problems of Consciousness* (Cambridge, MA: MIT Press).
- Van Gulick, R. (1995), 'What would count as explaining consciousness?', in *Conscious Experience*, ed. T. Metzinger (Exeter: Imprint Academic), pp. 61-79.
- Van Gulick, R. (2000), 'Inward and upward: reflection, introspection, and self-awareness', *Philosophical Topics*, 28, pp. 275-305.
- Van Gulick, R. (2004), 'Higher-order global states (HOGS): An alternative higher-order model

of consciousness', in Gennaro 2004.

Yablo, S. (1999), 'Concepts and Consciousness', *Philosophy and Phenomenological Research*, 59, pp. 455-63.