

Two HOTs To Handle

JENNIFER MATEY

Department of Philosophy
Florida International University

ABSTRACT:

Abstract: David Rosenthal's higher-order thought (HOT) hypothesis is one of the most widely argued for of the higher-order accounts of consciousness. I argue that Rosenthal vacillates between two models of the HOT theory. First, I argue that these models employ different concepts of 'state consciousness'; the two concepts each refer to mental state tokens, but in virtue of different properties. In one model, state consciousness is more consistent with how the term is typically used, both by philosophers and scientists, and in commonsense usage. This model, however, also has its problems. In the second part of the paper, I develop a modified version of the transitivity principle, thereby avoiding some complications that stem from the original transitivity principle. I suggest that Rosenthal occasionally employs this modified model himself, and that the inconsistency identified in the first section of this paper might really reflect Rosenthal's vacillation between these two versions of the transitivity principle. I offer one explanation for how this equivocation might have occurred. These two versions would result if articulations of the transitivity principle employed the term 'mental state' inconsistently, to refer on some occasions to mental state types, and on others to refer to a tokened mental state. I conclude by arguing, contrary to Rosenthal and others, that the theory is not incompatible with the view that conscious states are uniquely causal efficacious. Two HOTs To Handle: The Concept of State Consciousness in the Higher Order Thought Theory of Consciousness

1. Introduction

David Rosenthal's Higher-Order Thought (HOT) hypothesis is among the most widely argued for of the higher-order accounts of consciousness (Rosenthal 1991, 1993a, 1997 a,b, 2002 a,b,c). In this paper I address an inconsistency in his account of consciousness. Rosenthal's argument, I argue, vacillates between two independent models of the HOT theory. I suggest that at the heart of these two models are two different conceptions of state consciousness. In the first section of this paper, I review the two models and I refute several possible objections to my characterization of this as an actual inconsistency in the theory. In subsequent sections, I examine these models more closely. It is my contention that philosophers ought to invoke the principle of charity when reading critically. This means that where several models of a theory are put forth, philosophers should aim to assess the strongest among them. I argue that the second model is preferable to the first for several reasons. But that model is still problematic. Solving those problems would involve modifying one of the core features of the HOT theory- the Transitivity Principle.

In the second part of this paper I focus on the Transitivity Principle. First, I develop a modified Transitivity Principle. Then I hypothesize that Rosenthal occasionally employs this modified model himself. This slip may make sense of some problematic aspects of his theory. I proceed to suggest that the inconsistency identified in the first section of this paper might actually reflect these two versions of the Transitivity Principle. One version gives us a concept of state consciousness centered on a token target state, while the other only addresses mental state types. I offer one explanation of how this equivocation might have occurred. These two versions would result if articulations of the transitivity principle employed the term ‘mental state’ inconsistently, sometimes as a mental state type, and other times as a mental state token.

1.1 Rosenthal’s HOT theory.

In this section I make some preliminary distinctions and introduce the basic components of the Higher Order Thought theory.

The HOT theory is a proposal about the ‘what it’s like’¹ character of experience. Two of Rosenthal’s most basic claims are that *a) mental states are not intrinsically conscious* and *b) mental states become conscious only when they are accompanied by a higher-order-thought about them*. He defines a thought as a propositional attitude, an attitude such as a belief or a doubt, held toward a given content. A HOT is a thought about another thought, or as Rosenthal claims, a thought about a mental state. On his account, not all mental states are to be construed as thoughts; some mental states are purely qualitative.

Moreover, Rosenthal advocates a functional account of mental states that captures the particularly ‘mental’ components of the state, without assuming that phenomenality is a necessary feature of those states. These informational states are classified according to their general causal relationships rather than by their merely physiological ones.

A non-conscious mental state, Rosenthal argues, becomes conscious only if a certain relation holds between it and another state. The other state, the accompanying HOT, is distinct and yet contemporaneous with the mental state it is about. The higher order state is referred to as a transitive state since it is of or about something, whereas the lower order state it refers to is made to be intransitively conscious or ‘State Conscious’. The HOT is an intentional state with an assertoric attitude. Its content is a representation of the target state.

To advance his argument, Rosenthal appeals to common intuitions about the nature of conscious experience. He offers the intuitive correlation between thoughts we are able to have and sensory qualities that we are able to become aware of. Concepts, he argues, affect the way that we experience qualitative states. For example, in the case of wine tasting, higher order thoughts incorporate concepts that correspond to flavors recognizable as “dry”, “heavy” or “bright”. As we learn more concepts, we have more fine grained experiences and experience more qualitative complexities. By employing concepts, HOTs characterize conscious experience by weeding out and shaping mental qualities. Rosenthal argues that sensory qualities must surely be present on a non-conscious level, since they are determined by symmetries between qualities of objects and our physically determined sensory abilities². But the fact that we experience these qualities as being this way or that demonstrates that thoughts influence our conscious experience of those qualities. Rosenthal (2002c) explains:

being able to form intentional states about certain sensory qualities must somehow result in our being able to experience qualities consciously...The only plausible explanation is that a sensory quality's being conscious does actually consist in our having a HOT about that quality (p. 413)

He then goes on to explain further:

If one's HOTs couldn't classify one's sensations in terms of the sound of an oboe, but only that of some undifferentiated woodwind, having that sensation could not be for one like hearing an oboe. And if one also lacked any concept of the sound of a woodwind, what it would be like for one to have that sensation would then be correspondingly more generic..peeling away the weakest HOT would result, finally, in its no longer being like anything at all to have that sensation" (2002c, pp. 413-415).

To summarize, the HOT theory can be understood as accommodating for three phenomena. First, it provides an account of the distinction between conscious and non-conscious mental states. Second, assuming that consciousness isn't an intrinsic property of all mental states, it may be that consciousness is not a primitive and irreducible property. If this is true, then the HOT theory may bring us closer to finding a reductive account of consciousness by providing an articulated model. Third, the HOT theory provides a satisfying account of some phenomenally salient features of conscious experience such as the apparent fact that many of our phenomenal states are experienced as having a cognitive component.

1.2. An Inconsistency

In developing the HOT theory, Rosenthal puts forth a set of three not obviously compatible propositions:

1. HOTs always confer intransitive, or 'state consciousness', on the lower order states they are about.
2. HOTs can represent the states they are about accurately or inaccurately.
3. Target states do not become conscious unless the HOT's representational content includes informational properties that are type-equivalent to the content of the target state.

In this section, I will explicate Rosenthal's endorsement of the above statements.

1. HOTs always confer intransitive (state) consciousness on the lower order mental states they are about.

In *Thinking that One Thinks*, Rosenthal (1993b) writes: "A mental state is conscious... just in case one has a roughly contemporaneous thought to the effect that one is in that very mental state...having such a thought will *always* mean that the lower order thought it is about is conscious." (p. 199). This idea recurs throughout most of Rosenthal's work (1991, 1993a, 1997a,b, 2002a,b,c). In *Metacognition and Higher Order Thoughts*, Rosenthal claims that HOTs are of or about mental states, and that this 'of' is meant in an intentional sense. The relevant type

of intentional relation here is a two place relation. In particular, it is a relation between a higher order representing state and an independent target state. The target state becomes conscious whenever there is a thought that represents it. For example, Rosenthal writes; “If one re-describes to oneself a forbidden desire as a desire for something acceptable, the desire may well be conscious, even though not in respect of its actual content” (2000c, p. 234). The target desire becomes a conscious desire, despite the fact that its informational properties are not what we experience phenomenally.

The above quotes, and the recurrent statement of this idea in Rosenthal’s writings, establish a version of Rosenthal’s Transitivity Principle presented in the first claim: HOTs confer intransitive state consciousness on the lower order mental states they are about.

If the first claim is true, then it seems as though either the second or third should be false.

2. HOTs can represent states accurately or inaccurately.

In *Consciousness, Content and Metacognitive Judgments*, Rosenthal (2000a) writes, “Whatever the actual character of a mental state, that state, if conscious, is conscious in respect of whatever mental properties one’s HOT represents the state as having” (p. 213). Rosenthal reveals his commitment to this claim with examples he uses to persuade us that the HOT theory provides the best explanation for the phenomenal character of consciousness. These examples appeal to the HOT’s ability to characterize and thereby determine ‘what it’s like’ to be in mental states. His examples of intuitively plausible scenarios appeal for explanation to the influence of the higher order representation on the way we experience our conscious states. Some examples appeal directly to our intuitions about conscious experiences, while others appeal to our folk understandings of the way psychological phenomena distort the character of conscious states and misrepresent those states. In the next few paragraphs I will recount three of Rosenthal’s examples. One of the first kind, and two of the second

In 1.1, I presented two examples that Rosenthal uses where concepts are employed by HOTs, and determine what it’s like to have sensory qualities. Those examples came from wine tasting and experiencing musical sounds as being produced by distinct instruments. Another example that Rosenthal uses is that of walking through the woods, apparently aware of the branches that one steps over, but not qua branches, rather as part of an un-discriminated woodland environment. A hiker may have sensations of her environment, but those sensations are conscious as sensations of the un-discriminated scene. If she has more fine-grained thoughts about branches, trees, and leaves, then, Rosenthal insists, those thoughts are not conscious thoughts, but perhaps non-conscious thoughts that become conscious as part of a much less conceptually fine grained ‘woodland’ scene.

Another type of evidence used in support of the HOT theory comes from examples of misrepresentation. One criticism that has been used against the HOT theory concerns the possibility that HOTs might misrepresent the states we are actually in. If HOTs make first order states conscious by representing those states, then it seems possible that HOTs would on occasion misrepresent those states. Rosenthal turns this challenge around by making misrepresentation into an unexpected implication of the HOT theory which, he argues, reinforces the theory’s plausibility by contributing to its explanatory power. Any theory of consciousness,

he claims, must explain our folk understanding of conscious experience. Misrepresentation provides a technical framework for explaining what we already recognize as non-veridical conscious states and explain as such in our folk descriptions. Two examples of HOTs misrepresenting first order states frequently cited in Rosenthal's work are Freudian repression and the phenomenon of dental pain. In repression, Rosenthal (1986, 2000a, 2000b) hypothesizes, one unconscious desire is represented as something else.

Giving another example, Rosenthal writes:

Dental patients sometimes seem, from a first person point of view, to experience pain even when nerve damage or local anesthetic makes it indisputable that no such pain can be occurring. The usual hypothesis is the patient experiences fear along with vibration from the drill, and consciously reacts as though in pain (2002c, p.415).

In cases of repression and dental pain then, higher order states fail to accurately describe what is happening on a more fundamental psychological level.

I've summarized some of Rosenthal's arguments for the claim that a token target state becomes state-conscious by way of a higher order representation of that state. The HOT, as claim *two* implies, can characterize the target state by adding conceptual content, and it can represent that state either veridically or non-veridically, depending on the extent to which our experience of that state matches the content of the state itself. But this notion seems incompatible with the third proposition.

3. Target states do not become conscious unless the HOT's representational content includes informational properties that are type-equivalent to the content of the target state, i.e., unless the HOT's representational content matches the informational content of the target state.

This proposition is derived from a slightly different proposition that Rosenthal puts forth in 'Consciousness and Metacognition' (2000b). There he writes, "for a state to be conscious, one must be conscious of the state in virtue of its particular informational properties" (p. 267). I interpret this claim to mean that the HOT's representation of the informational contents of a mental state must contain content that is informationally equivalent to the particular content that the HOT is a representation of, in order for the target state to count as being intransitively conscious. I think that this is the most plausible interpretation, given the context in which this claim appears and the work that this idea is meant to do within that context. In what follows I will elaborate on that context and further substantiate my interpretation of that claim.

Rosenthal argues for this claim in a series of papers (2000 a,b,c) in which he makes a distinction between HOTs and other types of metacognitive judgments (MJs). All metacognitive thoughts, including HOTs, involve higher-order access to mental states we take ourselves to be in. An MJ, according to Rosenthal, tells us "whether or how likely it is that we know something" (2000a, p. 203). These judgments inform us about our cognitive condition and affect decision, action and subsequent cognitive processes (2000a). For Rosenthal, feeling of knowing judgments, of which the tip of the tongue phenomenon (TOT) serves as an exemplar, are prototypical instances of metacognitive judgments. In TOT, a person has a sense that some relevant information is present, but she is unable to recall that information. She is aware of being in the informational

state, although that informational state is not an intransitively conscious state. Rosenthal (2000a) claims that this disjunction between the way we are consciously aware of being in a mental state that is not intransitively conscious in TOT, and the way a HOT makes target states intransitively conscious, “helps close in on the exact way in which we’re conscious of our mental states when those states are conscious states” (p. 2004). So what is that way exactly?

Rosenthal goes on to state that the precise *way* in which we must be conscious of a state for that state to count as an intransitively conscious state, is *in respect of* the state’s specific informational content. Not only must we be conscious of what question that state provides the answer to, we “must be conscious of what the information is” (2000a, p. 205). Rosenthal argues, “A mental state’s being conscious consists not just in one’s being conscious of that state, but in one’s being conscious of that state under some relevant description” (2000b, p. 268). On the contrary, for example, “when I have George Orwell’s real name on the tip of my tongue but cannot recall it, I am conscious that I am in some state that carries the desired information, but I am not conscious of the state *in respect of* that information” (2000a, p. 205)³.

From the claims above it appears that Rosenthal makes clear the distinction between HOTs and other types of MJs by which we are ‘conscious of’ information that does not become ‘intransitively conscious’, with the argument that HOTs are higher-order judgments that represent states as having the particular informational properties of the states they represent. MJs, on the other hand, are states that do not contain, as a part of their content, informational contents equivalent to that of their targets.

Yet objections could still be made to the interpretation of this idea as put forth in statement three and therefore, to this as an actual inconsistency. In the section that follows I will refute these objections.

1.3. Objections

Objection One: Intransitive consciousness is a relational property of target states, not an intrinsic property.

One possible objection to my reading is that it reflects a misunderstanding of the term ‘state conscious’. State consciousness is not an intrinsic property of first order states, but rather a relational property. If consciousness did become intrinsic to the target state, then we would have reason to expect the way that state appears to consciousness to coincide with the informational properties of that state. Rather, Rosenthal (2000c) insists that while the target state is State Conscious, what matters to how the state appears is the manner in which it is represented.

The above objection cannot count as an objection to my reading. The inconsistency in the propositions remains despite the fact that state consciousness is not an intrinsic property of the target state. If statement two from the triad is correct, then the representation can be veridical or non-veridical. If three is correct, however, the state is conscious in virtue of the propositional or qualitative content it has, although again, not because consciousness becomes internal to the state and causes it to appear as itself, but rather because the mirroring of that state by the HOT allows us to consciously experience the target state’s informational contents.

Objection Two: Being conscious of a state ‘in respect of’ some property it has is a qualification on the represented property, but not a qualification on how that property is represented.

This objection is also directed at my interpretation of statement three. I claim that when Rosenthal specifies that conscious mental states must be conscious *in respect of* (2000a, p. 213) their informational content, he means they must be conscious *as* actually having those particular informational properties. I take Rosenthal’s use of *in respect of* as adding a qualification on the content of the conscious representation. But it could be alleged that he intended this as a weaker claim. In what follows I’ll canvas some weaker interpretations of statement three.

Weaker claim 1: Mental states are not conscious unless we are conscious of some aspect of their informational content, although the representation need not incorporate this information.

This interpretation of claim three limits the traits of the content that must be represented to the informational traits, for target states to count as being made to be state conscious, but it is nonspecific with respect to the aspects of those traits that must appear in the representation. This weaker rendering is problematic because it fails to accomplish what statement three was originally put forth to accomplish. Claim three is supposed to substantiate the distinction between HOTs and MJs. Both HOTs and MJs give us access to some characteristic of the target states’ informational content. MJs, after all, do represent aspects of the content of target states, such as what questions the information answers. These are functional properties of the state and Rosenthal’s informational account defines information in terms of its functional role. In addition, if no condition is put on the higher order representation, then we have no reason not to discount any old representation of those qualities as consciousness-conferring.

Weaker claim 2: The descriptive traits must be represented insofar as they are descriptions, rather than be represented insofar as they perform various functions, in order for them to be properly considered to become state conscious.

Yet, if this means that the claim limits only the qualities of the lower order state that are represented, but says nothing concerning the content of the representation, then this version doesn’t sufficiently explain the difference between metacognitive access and consciousness-conferring HOTs either. There is no reason to suppose that my thought that I am in a state with information that answers some question, is not actually a vague, or less than accurate representation of the information itself.

In addition, it seems natural to assume Rosenthal employs *in respect of* in identical ways in both statements two and three. We can read *in respect of* in statement two, as implicitly adding the qualification that the states are represented *in a way identical to* the informational contents represented. Recall that according to statement two, we are conscious of qualities *in respect of* the way the HOT represents the state as being. The HOT not only confers consciousness on the target state, it also provides the content of our experience of the target state. We are conscious of qualities in a way identical to the way the HOT represents the state as being. So if Rosenthal is using the phrase consistently, then statement three also adds the requirement that the HOT’s representation has informational contents that are identical to the target state it is about. And this would confirm the inconsistency.

Recall also that, “for a state to be conscious, one must be conscious of the state in virtue of its *particular* informational properties” (2000b, p. 267). If it is sufficient to make a state intransitively conscious that it be represented in terms of informational properties, but not as having a particular set of informational properties, then why does Rosenthal employ the qualifier ‘particular’? Moreover, if statement three puts a condition only on the class of characteristics of the target state to be represented, then what would the addition of the qualifier ‘particular’ add to that condition? Requiring that a state’s informational properties be represented, already implies that the particular informational properties of that state, rather than of another, be represented.

In summary, in this section I have shown that, if the HOT’s representation always makes the lower order state conscious as statement one suggests, then statement two adds to this the claim that HOTs can misrepresent targets, and thus, state consciousness can be conferred on a target state by a misrepresenting HOT. Statement three is inconsistent with this, since, as I’ve argued, it adds the requirement that representations be accurate reproductions of the actual content of the target in order for target states to become intransitively conscious. The principle of charity dictates that when a theory is ambiguous between two models, we assume the best model. In the next section I will determine which of the two models provides the best model for consciousness⁴.

2. The Two Models

Two models of Rosenthal’s theory can be constructed from the three claims above. I will characterize the difference between the two models as a difference in the concept of state consciousness employed by each. The first model is constructed from claims one and two. It states that target states become conscious states when they are represented, regardless of the content of the representing state. The second model is constructed from claims one and three. It states that target states become conscious states only if they are represented accurately. But note that in both of these versions of Rosenthal’s theory, it appears that state consciousness is a property of token target states. Even though the term refers to represented target states in both models, in each model the term refers to those targets in virtue of different properties of those target mental states.

2.1. State Consciousness in the First and Second Models

The concept of state consciousness that the first model employs refers to conscious states in virtue of their relational properties. Since, according to this model, HOTs can misrepresent target states and still make those targets conscious, the informational properties of the target state do not appear to be relevant to the state becoming state conscious. For ease, I’ll refer to this concept of state consciousness as ‘Relational State Consciousness’ (RSC). To be RSC, the intransitive state need merely bear a certain relationship to the HOT- the relation of being the object of that representation. No other properties of the target are necessary for the target state to become a state conscious mental state.

In contrast, the concept of state consciousness in the second model refers to conscious states in virtue of their informational properties in addition to their relational properties. Model two is constructed from statements one and three. It reads that target states become conscious just in case they are accurately represented by higher order states. According to this model then, in order to be state conscious, in addition to being represented by a higher order thought, a state must be the bearer of a particular set of informational properties- those particular informational properties must be equivalent to those that are reflected in the representational content of the HOT. In other words, the informational properties must be type-identical to properties that are part of the content of our phenomenal experience. For ease I'll call this concept of state consciousness 'Relational-Informational-State-Consciousness' (RISC).

But there are still two ways to develop a model of consciousness consistent with this second version. The first way is consistent with the rejection of statement two, and the unqualified endorsement of statements one and three. According to this model then, it would not be possible for HOTs to misrepresent their targets. Statement one gives us the claim that targets always become state conscious when they are represented by HOTs, and statement three gives us the claim that HOTs always represent conscious states accurately. So we might suppose that this model assumes that whatever the representational mechanism is that is responsible for HOTs, it simply does not allow for the possibility of misrepresentation.

The problem with this model, however, is that one of the common objections made to the HOT theory targets this very possibility for error (Byrne 1997). The objection runs like this: if HOTs are just a type of thought, then there must be at least some occurrences in which HOTs misrepresent their objects, since thoughts and perceptions in general are susceptible to these types of errors. The objection is based on the intuition that there is a logical possibility of error occurring. The objection moves from the observation that thoughts and perceptions are sometimes inaccurate, to the conceivability of HOTs making similar errors, to the positive possibility of error occurring. Whether or not HOTs can be erroneous, however, seems to be an empirical question. Until this empirical issue is resolved, I think that the most intellectually satisfying version of the HOT theory will be able to accommodate for the possibility of HOTs inaccurately portraying the target information, but the version will not necessarily assume that these misrepresentations actually occur⁵.

To account for this possibility, we might augment the individual claims. An easy solution is to put a stipulation on statement one: higher order representations might occasionally misrepresent their targets, but if this event were to occur the targets would not become state conscious, or would only become state conscious to the extent that they are represented accurately. The upshot here is that state consciousness will always be consistent with the informational model. Although misrepresentations can occur, misrepresented targets do not become intransitively conscious states.

The two models of the HOT theory I've presented appear to be architecturally and functionally equivalent. In each there is a target state and a representing state. In each the representing state is responsible for determining and providing the phenomenal content. And in each case it is the target state that is said to become state conscious. The core difference between them lies in what each model takes to be the essential properties of a conscious state. All else being equal, one reason we might favor one model over the other would be if one model were in better accord

with our common concept of what it means for a state to be conscious. Another reason to prefer one to the other would be if one model allowed us to make better sense of related phenomena. In the rest of this section I'll argue that the second (RISC) model is more satisfying in both respects.

2.2. Advantages of the Second Model

One problem with the purely relational concept of intransitive consciousness entailed by the first model is that it does not sit well with our intuitions about what it means for a state to be conscious. The commonsense or folk understanding of 'conscious state' associates these states with states whose informational content is equivalent to the information that is the content of phenomenal awareness. This is revealed by examples of the term being employed in common use. First, this concept of state consciousness is consistent with our uses of it in first person descriptions of the tip of the tongue phenomenon. The concept coheres well with the feeling of what is occurring from the perspective of the TOT subject who is trying to recall information she is conscious of knowing, and with the way that subject describes her experience. Recall that in TOT, the subject has the conscious experience of having a particular informational state without being able to produce the relevant information herself. In the next few paragraphs I develop this example.

On the one hand, it is arguable that in TOT experiences, a part of the informational content of the state one is trying to recall actually does become information that is phenomenally experienced. A set of functional relations that individuate that state are experienced. The set of informational properties that individuate a state might include such things as what question that state provides the answers to, and the feeling of knowing that is associated with actual knowing. So to the extent that this information is represented in the content of phenomenal consciousness, some of the state's informational properties become phenomenal content. Nevertheless, we intuitively judge that the target state is not conscious since the particular functional properties that we are seeking are precisely those that are unavailable⁶. Commonsense descriptions of the state we are seeking in the tip-of-the-tongue experiences characterize those states as not being intransitively conscious because the information we are seeking is not phenomenally experienced. This intuitive judgment can be taken as evidence that the informational properties of a state are relevant to our folk concept of intransitive consciousness.

The use of the term 'state conscious' to refer to states whose informational content is equivalent to the content of phenomenal awareness is also prevalent among researchers and philosophers who employ the term in a technical way. Because of this, the HOT theory is frequently misunderstood by philosophers. These critics gloss over the relevant relational properties of State Conscious states and import their presupposition that a state's informational properties alone determine whether the state should be considered conscious. These misunderstandings are manifest in the history of critical work directed at Rosenthal's theory. In the rest of this section I will provide several examples of these misunderstandings.

Ned Block (1995), Alvin Goldman (1993), and Fred Dretske (1995) allege that there is no good reason to think that being transitively conscious of a distinct state or object could result in a change to the properties of the intentional object. Their arguments run by analogy: if stones don't become conscious when we are conscious of them, and states of our body don't become

intransitively conscious when we are conscious of them, then why should mental states? This objection relies on the assumption that phenomenality becomes an intrinsic property of target conscious states. Rosenthal actually denies that phenomenality becomes an intrinsic property of the target state. But we can see how this assumption would be easy to have made; if one implicitly assumed the RISC concept of state consciousness, believing that a conscious state necessarily has informational content that is type-identical to the content of awareness, it is easy to make the further assumption that phenomenality is an intrinsic property of the conscious state. But this doesn't seem to be what Rosenthal is saying.

Consider also criticisms that resemble Karen Neander's (1998) critique of higher-order theories⁷. Neander notes that higher order theories account for conscious qualia through a two stage process. We receive qualitative information in the early states of visual processing, but this qualitative information only becomes conscious qualia when it itself is represented. Theories that involve representation allow for the possibility of misrepresentation, and Neander finds theories that allow for misrepresentation puzzling. If I represent an object as red, I have red quale. But if I then misrepresent to myself that red quale as green, I will experience a red quale, but 'what it's like' will be green- 'what it is like' for me to experience red, will be green. Neander notes that if this account is meant to explain qualia as traditionally understood- as the qualities of consciousness, then it is deeply puzzling. If qualia are by definition understood as being the qualitative properties of phenomenal experience, then what can it mean to say that qualia can become conscious without revealing their phenomenal properties- or as having different phenomenal properties? The problem with this criticism is that Neander doesn't accommodate for the unconventional way in which the term 'qualia' is employed in these theories. Likewise, if we interpret state consciousness in a way that matches the intuitive and commonsense notion of conscious states as states bearing some informational equivalence to the content of phenomenality, then we end up reading the HOT theory as implying some very unintuitive and counter-sensical things when the term applies to target states that are misrepresented. The information based concept of state consciousness employed in model two would not run into this sort of problem.

In summary, the two models seem to be architecturally and functionally identical. In each model a higher order state represents a lower order state. In each model, the HOT can misrepresent the target state. In each model it is the lower order state that becomes state conscious. And in each model the phenomenal content that an individual experiences is determined by the content of the higher order representation. It doesn't seem that any special metaphysical properties of the target state are picked out by either concept of state consciousness beyond the one's I have already identified. If this is true, then the only significant difference between the models lies in whether state consciousness is considered to be a property of all token target states just by virtue of their being represented, or whether state consciousness is a property that target states acquire to the extent that the phenomenal content of the representing state is equivalent to the target state's informational properties. If all else is equal, and the difference between the theories turns out to be merely a semantic or descriptive difference, then it is natural to prefer the model whose concept of state consciousness is most in line with the commonsense concept of intransitive consciousness. I've argued that when both the folk and philosophers and researchers say that some state is conscious, the term generally refers to states in virtue of their informational properties where those informational properties are type-identical to information that is phenomenally experienced.

In the next section I will show that although the second model is stronger, it also turns out to be problematic.

3. Additional Concerns

3.1 Problems for Target-State-Consciousness

Problems with the model that employs the Relational-Information concept of state consciousness become apparent when we consider cases of total misrepresentation of target states by HOTs. Georges Rey (2000) has suggested that the possibility of HOTs misrepresenting target states has negative implications for the HOT theory. In such cases there doesn't appear to be any state that we could properly call state conscious. If, as the second model suggests, token target states are only State Conscious to the extent that their informational properties are equivalent to the informational properties we experience phenomenally (as presented in higher-order representations), then target states whose informational contents are not in any way equivalent to the HOT's representation are not state conscious at all. What we seem pressed here to say is that although we have phenomenal experience, we are not in a conscious state. And this seems an undesirable consequence for any theory or descriptive model of consciousness.

An appeal to the first model won't help either. The first model may handle cases of misrepresentation; the concept of state consciousness there allows that misrepresented target states still become conscious states. Our experience of those states is merely inconsistent with the states' actual properties. But that model will run into problems similar to those that the RISC model runs into when the first model tries to handle a special category of misrepresentations that Rosenthal refers to as 'confabulations'. If HOTs are just a type of thought, then not only can they mis-present their objects, but they can also occur in the absence of their objects. Confabulations are hallucinatory HOTs. If state consciousness becomes a property of the mental item that the thought is intentionally directed toward, which I take to be the token target states they represent, then in cases of confabulation it seems that even though we have a representation of a non-existent state, no state of the individual actually becomes state conscious. That is to say that an individual has phenomenal experiences consistent with the higher order state's content- the notional state, without actually being in a conscious state. And yet it seems that any account that aims at the reduction of consciousness should include the claim that if an individual is having a phenomenal experience, some token mental state must be a conscious state.

In response to this problem, Rosenthal (2000c) claims that in cases of confabulation, HOTs are presenting 'notional' target states. They make such notional states into conscious states (2000c). But the idea of a notional target becoming state conscious seems even more perplexing. According to the version of the HOT theory presented, state consciousness becomes a property of token lower order states. A notional token target state is a target state that is not actual. So the question is, how can something that does not exist acquire a property?

3.2. The Transitivity Principle Revisited

There are a couple of options available to the HOT theorist. First, the HOT theorist could bite the bullet and concede that in cases of confabulation we have phenomenality without any token state becoming a state conscious state. Conceding this, however, is likely to alienate researchers who might otherwise have been sympathetic to the view; it is entirely counterintuitive to argue that phenomenality can be had without any token mental state being a conscious state. And if the architectural and functional model that the HOT theory gives us is a useful one, we should look for a better description to compliment the model. A second option would be to modify the Transitivity Principle- the first statement from the inconsistent triad.

I will offer a better description of the HOT theory. The description I offer will retain much of the wording of Rosenthal's original Transitivity Principle, but it will capitalize on ambiguities in some of the terms that are employed in the Transitivity Principle- the terms 'mental state' and 'conscious'. First, the modified version of the Transitivity Principle appeals to ambiguity in the term 'mental state', which can be used to refer both to mental state types and mental state tokens. The term 'mental state' refers to a mental state type if the user means by it a specific but abstract set of informational properties, and it refers to a mental token of that type if the user is referring to those properties instantiated at a particular time and/or place. The version of the Transitivity Principle that I have attributed to Rosenthal states that *token* target states acquire the property of state consciousness when we have higher order thoughts about them. By thoughts 'about' them, Rosenthal just means thoughts that represent those states. Thus, the original Transitivity Principle can be translated: *when higher order thoughts represent target states, the represented states become state conscious states and the representations provide the phenomenal properties and content*. This was roughly stated by the first proposition from the inconsistent set in section 1.2. The modification that I propose making to the Transitivity Principle gives us the following theory: to say that a represented 'state' becomes conscious when it is represented, is just to make a claim about mental state *types*, or abstract informational contents becoming conscious, instead of a claim about mental state tokens.

This modification also exploits a potential ambiguity in the term 'conscious'. Sometimes when we say that something is conscious we mean that some information has become 'the content of phenomenal experience'. But 'conscious' can also be used to refer to some particular token state becoming 'conscious' insofar as it acquires either the property of state consciousness, or of phenomenality. The original version of the Transitivity Principle seems to employ the term 'conscious' in this second more strict sense. The modification I am proposing uses the term 'conscious' just to make a claim about some information becoming the content of phenomenal experience, rather than a claim about a token state acquiring the property of consciousness. In summary, the modification I propose states the following: information, or mental state types, become conscious, just in the sense of becoming the phenomenal content I experience, when that information is accurately represented by a higher order thought. It thereby becomes the representational content of that higher order state. The HOT's informational properties are qualitatively identical to, although numerically distinct from, the target's informational properties. That information, insofar as it is conceived of abstractly as just information, becomes phenomenal content. But importantly, that information is only phenomenal content by virtue of becoming part of the representational properties of the higher order state. According to this version, although we might say that the target's informational contents become 'conscious', this should be understood in a loose way, since it is always a state token that becomes conscious in

the second more strict sense, by acquiring the property of either state consciousness, or of phenomenality.

One consequence of utilizing this modified version of the Transitivity Principle would be that any target informational properties that are incorrectly represented, cannot be said to be, even loosely put, conscious or phenomenal content. To adopt this modified Transitivity Principle, the notion that misrepresented states are made conscious when they are misrepresented would have to be abandoned. This is a natural consequence of this modified theory since this theory is not about token target states acquiring the property of state consciousness. Rather, this modified version is about unconscious or non-conscious information that is present in lower order states, becoming phenomenal content by being represented by a higher order state which consists of a belief about a representational content that is generally type-identical to the target state.

One might object that this modified version of the Transitivity Principle would fall short of a theory of state consciousness because state consciousness is a property of token states, not of mere abstract informational contents. So if some piece of information can loosely be said to be conscious, or experienced phenomenally, it is only because some token of that information is state conscious. One way to avoid this superficial problem along with the problems that arose for the description that made state consciousness a property of a token target state, would be to include the claim that a token state is state conscious just in case it is the bearer of phenomenal properties. This seems reasonable and it would have the consequence of making the representational component of the token higher order state the state conscious token. This model has all of the advantages that the model based on the RISC concept of state consciousness; the concept fits even better with our commonsense concept of state consciousness which refers to states whose informational content is equivalent to the information that we experience phenomenally, and it is consistent with the way that the term is used by researchers and philosophers. In addition, this modified description of the HOT theory retains the functional organization and architectural components of the original models. Only the description of that architecture has changed. But this revised description is logically more coherent than the description based on the target centered Transitivity Principle, and it comes at no real metaphysical cost.

A superficial obstacle to a version that would make state consciousness a property of the higher order state is that part of the theory will end up running into problems fitting with other aspects of the HOT theory⁸. In particular, this model seems inconsistent with Rosenthal's contention that HOTs themselves are not State Conscious unless they are represented by second order HOTs. But the modified version can easily accommodate for this problem. For example, keeping consistent with the modified version whereby we used the terms 'mental state' to refer to informational contents, and the phrase 'become conscious' to mean 'become the content we phenomenally experience', we might instead say that the full informational content of a HOT is not explicitly conscious, it is not presented in phenomenal experience unless the HOT is meta-represented. So while the HOT is state conscious, its full content is not made explicit in the content of phenomenal experience. Recall that the full content of the higher order state consists not only of the representation of the target, but also of a belief held toward that representation. The propositional content is conscious, and that content is also implicitly endorsed. The assertoric attitude is conscious as an implicit endorsement of the content of the state it is about, but it is not conscious as an explicit belief. The full content (belief + representational content)

only becomes explicit when HOTs themselves are represented in even higher order thoughts. Those introspective states bear representations of beliefs about proposition-like contents. The explicit phenomenal experience of being in an introspective state might be something of the sort, “I believe that the grass is green”, and this explicit representational content is in turn, also implicitly endorsed.

Finally, I think that it can be argued that in some of his work, Rosenthal is implicitly advocating for something like the alternative model of the Transitivity Principle presented here. For example, in his response to Georges Rey, Rosenthal (2000c) might be appealing to a model like this when he claims that in cases where we have hallucinatory HOTs the states that become conscious are notional states. The most charitable way to construe Rosenthal’s claim that notional states are made conscious, is to attribute to Rosenthal the modified version of the Transitivity Principle. In elaborating on his claim, Rosenthal (2000c) likens hallucinatory HOTs to events such as thoughts about non-existent entities like Santa Claus (2000c). He claims that HOTs are intentional states that posit a relation between us and something that the thoughts are about. Following up on these two claims, the claim about notional states becoming conscious might translate to something like this: *a higher order state makes ‘conscious’ or makes into ‘phenomenal content’, an informational content or mental-type that purports to be about, represent, or to refer to, a state that doesn’t in fact exist.* By a ‘notional state’, Rosenthal may just be referring to a mental-type that is not present in any lower order state but that it is about it in the sense of referring to that state. According to this modified version, to say that the notional mental type becomes conscious, is just to say that it becomes phenomenal content. According to the previous interpretation of this claim, Rosenthal seemed to be suggesting here that non-existent states could acquire the property of consciousness. This is mysterious. On the other hand, Rosenthal’s claim about notional states is logically coherent and straightforward if we read it in light of the modified version of the Transitivity Principle.

But if we take Rosenthal’s claim about notional states becoming conscious states as reflecting his endorsement of the modified version of the Transitivity Principle that I’ve presented, then it is not clear how his employment of this model can be reconciled with the models of the HOT theory based on the version of the Transitivity Principle that I identified in the first sections of this paper. There, the first proposition from the inconsistent set along with the second, suggested fairly straightforwardly that the theory is about state consciousness becoming a property of represented state tokens.

I think sense can be made of this apparent conflict as well. The existence of these two versions of the Transitivity Principle can be read as a consequence of not being consistent or specific with respect to whether the use of the term ‘mental state’ is meant to refer to mental state types or mental state tokens. Recall that I constructed the modified version of the Transitivity Principle by exploiting ambiguities in the terms ‘mental state’ and ‘conscious state’. If Rosenthal had made such an equivocation himself, it may have resulted in one model of the transitivity principle based on state tokens, and one model based on state types. One model would result in a theory of state consciousness in which token target states become state conscious when they are represented: *a state is state conscious only to the extent that it is represented by a HOT.* The other model of the Transitivity Principle would result in a claim about the transference of information or type properties between non-conscious states that we are conscious of, and

conscious states with phenomenal properties: *a mental state-type becomes phenomenal content when we are conscious 'of', by representing, a token state with type identical contents.*

Finally, returning to the argument I presented in the first section of this paper, and the incompatible set there, I'll now make a final speculation concerning the relevant propositions and the inconsistency they seem to suggest.

Perhaps that inconsistency actually reflects the two different versions of the Transitivity Principle rather than two different target centered notions of state consciousness. We can still read statements one and two as supporting a model based on the target centered concept of state consciousness. The concept of state consciousness employed there is the relational concept. That model employs the first version of the Transitivity Principle: a state conscious state is a lower order state represented by a higher order state. The second model on the other hand would deny statement one as it is stated, but accept a modification of that statement that reflects the modified version of the Transitivity Principle based on the alternative meanings of 'mental state' and 'conscious'. Mental state types that are instantiated in non-conscious states, become phenomenal content when they are represented in higher order states. That is how that information becomes information that is present in a state with phenomenal properties.

I suggest that statement three might be taken as evidence for Rosenthal's endorsement of this modified version of the Transitivity Principle, rather than as support for the information based model of state consciousness that refers to lower order target states that I had identified in the second section of this paper. Information from target states becomes conscious (i.e. phenomenal content) only if that information is represented accurately. If it is not represented accurately, then some other information becomes phenomenal content, but not the information that is represented.

I leave it to the reader to decide whether this modified version of the transitivity principle, along with this alternative notion of state consciousness, retains enough of the original proposal in order to inherit that proposal's force. Functionally and architecturally, the two models seem identical. The difference is in the descriptions of the models. The modified proposal suggests that information abstractly conceived, that is entertained on a non-conscious level, becomes phenomenal content, or loosely put 'conscious content', when it is part of the representational content of particular kinds of states. Those are states that purport to represent the contents of other states, and perhaps they are generally caused by the states they purport to be about. They are special kinds of states capable of bearing phenomenal properties. But this model doesn't explain how the token higher order mental states that bear those contents acquire their phenomenal properties; it does not explain why or how those token states became State Conscious states. That would require a different kind of theory.

The present theory seems to work well as a model that accounts for the difference between non-conscious and conscious states. It provides an articulated account of consciousness, and gives an account of the informational properties conscious states have, the propositional form of that content, and how information that we entertain non-consciously becomes part of the phenomenal content we experience. It also accounts for the cognitive component of many phenomenal experiences such as the way that qualitative properties seem affected by concepts we learn and are able to employ. Those were some of the original objectives of the HOT theory, in addition to providing an architectural and functional model that might be useful for consciousness research

in other disciplines. Reductionists, like Rosenthal, often argue that those disciplines that provide an account of how physical or functional states are capable of having phenomenal properties are the disciplines that are most likely to make gains in solving the hard problem.

3.3. Additional Concerns for the HOT Theory

One further question for the HOT theory considers whether HOTs can add conceptual content, characterizing represented content, and still be said to be making the information that is represented conscious? The alternative would be to consider this characterized representation different informational content from the information represented.

The relationship between the first order content and the content of the higher-order state can be described along qualitative and quantitative dimensions. I call 'qualitative' equivalence the equivalence between the meanings and senses of the actual intentional contents. To make the informational properties that target states contain into phenomenal properties of HOTs requires that at least part of the HOT's representation be of first order states in respect of, or in a way qualitatively equivalent to, the content of the first order state. I call 'quantitative' equivalence the equivalence between parts and wholes of intentional content. I propose that if HOTs do make first order contents into phenomenal contents in terms of representational characterizations of those states, then they must do so in a way that would respect the part-whole relationship between 'senses' and their conceptual parts. The higher-order representation of a lower order property must be equivalent to the target content, and also be a logical part of the HOT's total informational content, in order for that informational content to count as equivalent to the phenomenal content the individual experiences.

According to this model then, despite the fact that just one content becomes conscious, there is typically still a repetition of the informational content in a token target state, and in the higher-order token's representation of that state. The states are one state insofar as we refer to the states in a restricted sense, as merely psychological states- informational or functional contents. But they are two states insofar as we refer to those identical contents as being manifest in different forms, physically realized in two locations or configurations, or differently encoded. This doubling up of content may make good sense if the higher-order state is necessary in order to provide a conversion of the target state, whose informational content may be encoded in a way that is functionally useful but different from the language accessible to consciousness, or if the neural underpinning of that state is not a consciousness producing correlate. The higher order state may provide a necessary conversion of that content into what-it-is-like to have this particular experience.

3.4 Consciousness and Causality

In closing, I address one more problem posed by critics of the HOT theory. This is the concern about the causal efficacy of conscious states.

On the intuitive level it seems that there is something about our phenomenal experiences that is essential to the causal roles of those states. But the HOT theory doesn't seem to afford conscious

states any additional causal powers that would be due to their phenomenal properties. The lack of causal efficacy of conscious states has been one of the major criticisms of the HOT theory. A familiar version of this criticism comes from Fred Dretske. Rosenthal writes (2002c),

Dretske has noted that theories on which a state's being conscious consists in one's being transitively conscious of the state, seem unable to explain how a mental state's being conscious could have any function. Being transitively conscious of a state, on these theories, makes no difference to the state's non-relational properties. So the state's being conscious will make no difference to its causal role nor, therefore, to its function. (p. 416)

Rosenthal himself then goes on to deflate our intuitions concerning the causal role of conscious states (2002c),

It's easy to overestimate the degree to which a state's being conscious does actually play any role. It's inviting to think, for example, that a state's being conscious somehow enhances any planning or reasoning in which that state figures. But the role a state plays in planning and reasoning is due to the content the state has, and that content will be invariant whether or not the state is conscious. So whether or not a state is conscious will not affect the state's role in planning and reasoning. We find it tempting to insist that a state's being conscious affects planning and reasoning when we consider actual cases in which the planning and reasoning are conscious. But those cases tell us nothing unless we compare them to non-conscious cases, to which we have no first-person access. Intuitions cannot help here. (p. 416)

The target centered model of the Transitivity Principle has a difficult time allowing for the causal efficacy of phenomenal states. If consciousness is a relational property of conscious states, how could acquiring the property of consciousness make any causal difference for state conscious states? Relational properties do not generally influence or enhance a state's causal powers.

But the modified version of the Transitivity Principle may seem open to a similar objection. While intuitively, it seems that phenomenality should play a role in determining the causal properties and powers of a mental state, if the HOT's representation of the target is informationally equivalent to the target information, what difference would the HOT's contribution of phenomenal properties make to the already causally efficacious lower order state? Lower order informational content is sufficient to explain mental causation. The doubling of that informational content on a conscious level wouldn't appear to add anything to the information's causal powers.

Consider this example. Suppose that I have the unconscious thought that the cat is on the counter. That unconscious thought, together with some desire, seems sufficient to cause me to remove the cat from the counter. If that thought is important enough and not for some reason blocked by protective mechanisms from becoming conscious, the thought might become a conscious thought. But the thought qua thought would have identical causal effects. The conscious thought that the cat is on the counter, has the same informational properties that the unconscious target does. Information itself, aside from its position along a conscious/non-conscious dimension, doesn't seem to be the sort of thing that can vary in intensity- at least not the sort of information expressed in a phrase such as, "the cat is on the counter".

I don't think it is necessary to draw the conclusion above. Not all information is emotionally or affectively neutral as the phrase "the cat is on the counter" appears to be. Moreover, even mental states that appear to be without affective content when they are analyzed semantically, are frequently accompanied by an emotional valence nonetheless. The functional associations of a state include the state's link to our varying wishes, hopes, fears, desires, along with emotions such as joy and sadness.

I can have a slight thirst, or I can have a strong thirst. Likewise, I can have a bit of a desire to drink some water, or I can have a strong desire for some water. While the informational content of a mental state (that I would like some water) either exists or does not exist, the desire itself can vary in intensity. For the sake of explication, let's suppose that you could rate a desire in accord with its level of intensity. My desire for water at a given time, is designated a rating of '4' on a 1-10 dimensional scale. When my desire becomes conscious via a second state that represents that target desire, I then have one conscious informational state, but two internal states with that informational content and each with individual scores of '4'. The doubling of that content makes no difference to the information itself, but insofar as that information is also related to emotional states which *can* vary in intensity, the doubling of the state might result in a doubling of emotional intensity. I now have a desire for water with a sum emotional intensity contributed from both states- an '8'. An '8' might affect my behavior in a way that a '4' would not. The semantic account of the content of that state would miss the affective dimension and make it seem that the conscious state adds nothing to the information's causal powers.

Consider again the apparently neutral thought "the cat is on the counter". Associated with this thought are connections to an indeterminate amount of further mental states with their own emotional valences. So even in cases where the mental state in question does not obviously contain an emotional charge, a rather complicated set of emotional charges accompany the state in virtue of its connections to other thoughts and states. Moreover, these connections contribute to the content of the functional state in question. The repetition of "the cat is on the counter" in virtue of a HOT entails the repetition of all the connections entailed by the state. Given the conclusions concerning intensity from the above paragraph, even not-so-obviously affective informational contents are likely differently efficacious in virtue of higher-order conscious representations of them.

Thus, even if the causal properties of conscious states don't occur in virtue of those properties being phenomenal properties, the additional causal properties that consciousness contributes to mind could still occur in virtue of consciousness qua a special kind of mental state. Insofar as consciousness is generally produced by the interaction between a token target state and a higher-order representation of that target, the higher-order state is capable of intensifying the emotional or affective components of states that factor in their ability to influence and cause other states and behaviors on both non-conscious and conscious levels.

3.5 Conclusion

In the first sections of this paper I reconstructed two versions of Rosenthal's Higher Order Thought theory. I've argued that we can frame this distinction by saying that the first model's

concept of state consciousness is concerned merely with the state's relational properties, whereas we can develop the second model in a way that takes bearing a specific set of informational properties to be relevant as well. I've also argued that if the models are functionally the same, then the second model is preferable. One reason is that it fits better with the commonsense concept of intransitive consciousness. All else being equal, a theory that fits well with this understanding, better accounts for our experiences and should be preferred. In addition, some thoughtful objections to the Higher Order Thought theory may have been misled by the unintuitive concept of state consciousness in the first model. But there are still problems with the second model. The second model leads to the unintuitive consequence that we may have phenomenal experiences without having conscious states. I've highlighted this problem and suggested that it might be averted by modifying Rosenthal's Transitivity Principle. My modification to the Transitivity Principle requires exploiting ambiguities in the terms from statement one of the incompatible set I presented in the first section of this paper, and it would seem also to require making additional alterations to the HOT theory. There is some evidence that Rosenthal himself employs the modified version I suggest. In fact, the inconsistency that I have identified in the beginning section of this paper can be recast in terms of these two versions of the transitivity principle. If we accept the modified Higher Order Thought theory as offering a viable hypothesis about the nature of consciousness, I've suggested that the theory also withstands another objection often made against it; it is not incompatible with the causal efficacy of conscious states. Although conscious states may not make a causal difference merely in terms of their phenomenal properties, I've offered an alternative suggestion as to how they can still be causally efficacious. HOTs add to the causal status of informational content we already entertain on a non-conscious level by intensifying the mental-type's affective dimensions.

The author would like to thank Jayanta Hegde, David Patten, David Rosenthal, Harvey Cormier, Uriah Kriegel and Patrick Grim, for helpful comments on an earlier draft of this paper. The paper was also improved in light of discussions with the Consciousness Discussion Group at the University of Arizona's Center for Consciousness Studies, and with the Philosophy of Mind reading group also at the University of Arizona. Special thanks to Mike Bruno for suggesting the first half of the current title.

Notes:

¹ I use this expression interchangeably with the term 'phenomenality'. Some use 'phenomenology' to refer to subjective qualitative experience but I choose to reserve 'phenomenology' for the philosophical movement which goes by the same term and for the particular style or method associated with that movement.

² The idea here is that there is a consistent correspondence between perceptible properties of objects and phenomenal properties of perceptions. See Rosenthal: 1999a, 1999b, 2001a.

³ A reviewer suggests that metacognitive judgments such as tip-of-the-tongue states be read as states in which the desired information is conscious information, but not in respect of the informational contents we are seeking. I agree that this seems like a tenable interpretation of the TOT phenomenon according to the constraints of the HOT theory. For example, the information we are seeking may be conscious, but it is represented as information we cannot recall. But this is not what I took Rosenthal to be saying. I take him to drawing a distinction between TOT states and HOTs, rather than arguing that TOT states are an example of a unique type of HOT.

⁴ I will sometimes refer to the two models that follow from these propositions as being incompatible with each other. This is not, strictly speaking, a logical incompatibility. I assume, however, that a theory should not rely on ambiguous terms. An undesirable consequence of this ambiguity would be the endorsement of two models which

each employ the same term differently. It is to the extent that we want terms employed consistently that the propositions are inconsistent.

⁵ It might be argued that evidence for cases of misrepresentation are plentiful. But on the other hand, cases that appear on the intuitive level to be misrepresentations or confabulations, may not actually be. It may be that on some occasions, the anticipated target is somehow blocked from becoming the content of a HOT. On the lower level, the anticipated target might causally implicate the formation of an alternative, more acceptable lower order target state, and the HOT is redirected to the alternative state. Our folk descriptions might then mis-label these states as 'misrepresentations'.

⁶ Rosenthal (2000c) offers this explanation of our intuitive understanding of TOT in My claim here is just that our intuitive understanding of this phenomenon coheres well with an explanation constructed from the second model.

⁷ Neander's paper is directed at Lycan's higher order perception account, but she implies that it can be extended to higher order theories more generally.

⁸ Rosenthal claims that although the HOT determines the character of phenomenal experience, the HOT itself is not a conscious state unless it becomes the target of an even higher-order state. This only occurs, according to Rosenthal, in instances of introspection. It seems right that the HOT theory should explain introspection by positing second-order HOTs about HOTs, which turn the full informational contents of target HOTs into phenomenal content. But according to the model suggested here, the representational contents of the first order HOTs do become phenomenal properties even before the second order HOT is directed at the state. The HOT then is itself at least partly state conscious. It could still be the case however, that the belief component of the higher order thought that concerns the representational content remains implicit until the HOT itself is meta-represented. With first order HOTs we may have some of the phenomenal experiences that are associated with believing, we accept that information by default and we are ready to act on the information. But the belief is none-the-less not explicit; so we are not aware of it insofar as it is an act of believing and it is not something about which we are likely to make verbal reports. If the HOT were to be represented in an even higher order state, as in states of introspection, the implicit belief might then become explicit; we would become conscious of the belief qua a belief. For this to happen, the belief might have to be semantically encoded by the second order HOT. In any case, the idea that higher order states are not state conscious would have to be revised. For the modified transitivity principle, it seems sufficient to say that the full informational contents of a higher order thought, do not become the phenomenal content we experience unless a second order HOT contains a representational content that is type-identical to the informational properties of the target HOT.

References

Block, Ned. (1995) "On a Confusion about a function of Consciousness." *Behavioral and Brain Sciences* 18 (2), 227-247.

Byrne, Alex. (1997) "Some Like it HOT." *Philosophical Studies* 86: 103-129.

Dretske, Fred. (1995) *Naturalizing the Mind*. Cambridge, MA: MIT Press/Bradford.

Goldman, Alvin I. (1993). Consciousness, Folk Psychology and Cognitive Science. *Consciousness and Cognition* 2 (4), 364- 382.

Lycan, W. (1987). *Consciousness*, Cambridge, MA: MIT Press.

Lycan, W. (1996). *Consciousness and Experience*, Cambridge, MA: MIT Press.

Lycan, W. (2001). A simple argument for higher-order representation, *Analysis*, 61, 3-4.

Lycan, W. (2004) The superiority of HOP to HOT. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 93-114). Amsterdam: John Benjamins.

Neander, Karen. (1998) "The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness." *Philosophical Perspectives, 12 Language Mind and Ontology*. 411-434.

Rey, Georges. (2000) "Role, Not Content: Comments on David Rosenthal's 'Consciousness, Content, and Metacognitive Judgments'". *Consciousness and Cognition* 9, 224-230.

Rosenthal, David M. (1986a). "Two Concepts of Consciousness". *Philosophical Studies* 49, 3 (May): 329-259

Rosenthal, David M. (1991a). "The Independence of Consciousness and Sensory Quality". In *Consciousness: Philosophical Issues*, 1, ed. Enrique Villaneuva. Atascadero California: Ridgeview Publishing Company, pp 15-36.

Rosenthal, David M. (1993a). "State Consciousness and Transitive Consciousness". *Consciousness and Cognition*. 2, 4 (December): 355-363.

Rosenthal, David. (1993b) "Thinking that One Thinks," in *Consciousness: Psychological and Philosophical Essays*, ed. Martin Davies and Glyn W. Humphreys, Oxford: Basil Blackwell, pp. 197-223;

Rosenthal, David M. (1997a). "A Theory of Consciousness." In *The Nature of Consciousness: Philosophical Debates*, ed. Ned Block, Owen Flanagan and Guven Guzeldere. Cambridge, MA: MIT Press, 729-753.

Rosenthal, David M. (1997b). "Phenomenal Consciousness and What it's Like." In *Behavioral and Brain Sciences*, 20, 1 (March): 64-65.

Rosenthal, David. (1999a) "The Colors and Shapes of Visual Experiences," in *Consciousness and Intentionality: Models and Modalities of Attribution*, ed. Denis Fisette, Dordrecht: Kluwer Academic Publishers, pp. 95-118

Rosenthal, David M. (1999b) "Sensory Quality and the Relocation Story," *Philosophical Topics* 26, 1 and 2 (a special issue in honor of Sydney Shoemaker), pp. 321-350.

Rosenthal, David. (2000a) "Consciousness, Content, and Metacognitive Judgments" *Consciousness and Cognition*, IX, 2, Part 1

Rosenthal, David. (2000b) "Consciousness and Metacognition", in *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*, ed. Daniel Sperber, New York: Oxford University Press

Rosenthal, David. (2000c) "Metacognition and Higher-Order Thoughts" (extended response to commentaries), *Consciousness and Cognition*, IX, 2, Part 1

Rosenthal, David. (2001a) "Color, Mental Location, and the Visual Field", *Consciousness and Cognition*, IX, 4: 85-93.

Rosenthal, David M. (2002a) "Consciousness and the Mind", *Iyyun, The Jerusalem Philosophical Quarterly* 51, 3: 227-251.

Rosenthal, David M. (2002b). "How Many Kinds of Consciousness?" in *Consciousness and Cognition*. 11. 653-665.

Rosenthal, David M. (2002c). "Explaining Consciousness". In *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David Chalmers. Oxford: Oxford University Press. 406-421.