This is a draft of a paper that will appear in corrected form in *Philosophical Psychology*. Please contact the author for the final draft before citing.

Reduction and the Determination of Phenomenal Character

Jennifer Matey  
Department of Philosophy  
Florida International University

**Abstract:** A central task of philosophy of mind in recent decades has been to come up with a comprehensive account of the mind that is consistent with materialism. To this end, philosophers have offered useful reductive accounts of mentality in terms that are ultimately explainable by neurobiology. Although these accounts have been useful for explaining some psychological states, one feature, phenomenality or consciousness, has proven to be particularly intractable. The Higher-Order Thought theory (HOT) has been offered as one reductive theory of consciousness. According to HOT, a mental state is conscious if it becomes the content of a suitable higher-order thought that one is in that mental state. In recent years, critics have lodged a series of challenging objections to the view and several alternative theories have been proposed in response to these objections. This paper offers a defense of the traditional Higher-Order Thought theory. First, two different models of consciousness based on HOT are distinguished. The paper argues that one of the models is better supported by the HOT literature. It is then demonstrated that the better supported model is not vulnerable to the objections most commonly lodged against the HOT theory. Finally, it is shown that alternative self-representational theories do not improve upon the HOT theory in the way that they are proposed to. In fact, each of the alternative self-representational views reviewed here is vulnerable to a unique set of problems. In light of these factors, HOT still offers a viable reductive solution to the hard problem.

**Introduction**

A central task of philosophy of mind in recent decades has been to come up with a comprehensive account of the mind that is consistent with materialism. To this end, philosophers have offered useful reductive accounts of mentality in terms that are ultimately explainable by neurobiology. Although these accounts have been useful for explaining some psychological states, one feature, phenomenality or consciousness, has proven to be particularly intractable. The Higher-Order Thought theory (HOT) has been offered as one reductive theory of consciousness. According to HOT, a mental state is conscious if it becomes the content of a suitable higher-order thought that one is in that mental state. In recent years, critics have lodged a series of challenging objections to the view and several alternative theories have been proposed in response to these objections. This paper offers a defense of the traditional Higher-Order Thought theory. First, two
different models of consciousness based on HOT are distinguished. The paper argues that one of the models is better supported by the HOT literature. It is then demonstrated that the better supported model is not vulnerable to the objections most commonly lodged against the HOT theory. Finally, it is shown that alternative self-representational theories do not improve upon the HOT theory in the way that they are proposed to. In fact, each of the alternative self-representational views reviewed here is vulnerable to a unique set of problems. In light of these factors, HOT still offers a viable reductive solution to the hard problem.

1. Consciousness as higher-order representation

‘Consciousness’ refers to the unique subjective feeling that many of our mental states have. Another way to describe this feature of mentality is to invoke the notion of ‘what it is like-ness’. Whenever someone has a conscious mental state or a phenomenal experience, there is something that it is like for them to be having that mental state (Nagel 1974). A complete reductive account of the mind must account, not only for thoughts like beliefs and desires, but it must also be able to explain why or how there should be something that it is like for you to be in a conscious mental state. This challenge has been termed the ‘hard problem’ of consciousness (Chalmers 1996).

It does not appear that the hard problem can be easily resolved by appeal to standard scientific methods. Whereas we can explain most cognitive processes by providing a third person account of the mechanisms responsible, consciousness seems to be irreducibly subjective (Nagel 1974). This has led some to suspect that there must be a gap, if not in nature, at least in our ability to conceptualize this aspect of the natural world (Levine 1983, McGinn 1989, 1991).

The Higher-Order Thought theory was originally put forward by David Rosenthal as a reductive solution to the hard problem. A reductive solution must show that the set of facts to be reduced is actually constituted by another set of cognitive or material facts, without remainder. In the case of consciousness, we are looking for a set of cognitive or physical mechanisms which, when present, will entail consciousness. Cognitive mechanisms are generally easy to account for in terms of material properties.

A conscious state, according to HOT, is a state that plays a particular role. A theory of consciousness should describe the cognitive mechanisms that play that role. According to the HOT theorist, intuitively, having a conscious state is the same thing as having a state that we are aware of ourselves as being in. Note that one way that we describe a non-conscious mental state is to say that it is a state that we are completely unaware of being in. The difference between conscious and non-conscious states seems to consist just in the fact that conscious states are states that we are aware of being in. It is common to hold that we become aware of things by representing them with mental states such as thoughts. The HOT hypothesis fits with this common view about awareness by positing this to be the way we become aware of our own mental states, by representing them in a thought.
The upshot then is that consciousness reduces to the presence of a suitable higher-order thought describing oneself as being in the mental state that it represents. The representing state is called ‘higher-order’ because it is a meta-state that describes oneself as being aware of being in another mental state. The state we are aware of is called the lower order state. Sometimes it is referred to as the target.

In summary, the higher-order thought theory explains the intuitively appealing idea that conscious states are states the subjects of consciousness are aware of, or conscious of being in. This notion is sometimes referred to as the transitivity principle (TP) (2000b 2005). The HOT theory explains this intuitively appealing idea by making a mental state’s being conscious consist in it being represented by a suitable higher-order thought. The theory is appealing insofar as it would provide a genuine reductive account of consciousness. But this alone is not sufficient to motivate the theory. Independent support is required if HOT is to be considered a plausible theory about the nature of consciousness. The section to follow discusses some of the evidence that has been provided for the view that consciousness consists in instantiating the appropriate type of meta-cognition.

2. Making the case for the higher-order thought theory

The HOT theory is most notably defended by David Rosenthal who has supported the theory by an argument of inference to the best explanation. He presents several aspects of phenomenal character that seem to require explanation and then proposes that the HOT theory provides a satisfying explanatory framework in the case of each of these phenomena. This section reviews three facets of subjective experience that, according to Rosenthal, the HOT theory explains.

First, it is observed that something beyond our sensory abilities can influence the subjective experience of our sensory qualities when we are conscious of those qualities. Our physically constrained sensing abilities determine which sensible qualities of objects we are capable of experiencing. These faculties provide us with sensory information that is isomorphic to sensible properties of objects. Since we know our sensory faculties to be highly attuned, it is expected that we would perceive scenes as having a great deal of qualitative complexity. Yet, our phenomenal presentation of the world is not always experienced as being as qualitatively rich as our sensory abilities should allow for. On the basis of this observation, Rosenthal (2002) has hypothesized that higher-order representations of non-conscious qualitative states employ concepts that filter, characterize, and otherwise lend form to our qualitative experience of the world.

Rosenthal illustrates this point by appealing to the example of perceiving a wooded lot (2002). On different occasions and in different contexts the lot may be experienced in more or less detail. For example, when walking alone, one might experience the scene in great detail. But on a different occasion, such as when the woodland is serving as background to an engrossing conversation, the scene might be experienced in sparse detail. This would make sense if the way that the scene were experienced could be
influenced by the type of concepts about the scene one deploys in a higher-order thought (2002). When alone, we are capable of being very attentive and may have the resources available to entertain detailed concepts about the scene. But if we are engaged in conversation, we may have fewer available cognitive resources and use only vague concepts. It seems reasonable to hypothesize that whether one experiences individual leaves and branches as leaves and branches or merely as an indistinguishable aspect of the woodland, coincides with the structure resulting from the concepts that are applied in higher-order thoughts. Thus, the HOT theory accounts well for the fact that we may be more or less aware of qualitative complexity on different occasions even if our capacity to sense remains constant.

Second, the sensory qualities that we do experience can be experienced in particular ways, as being this way or that way. The fact that we experience qualities as characterized in particular ways conforms with the view that the way we are aware of qualities can be influenced by something over and above our physically determined sensing capacities. For example, it often seems that learning new concepts and words for things influences the type of experiences we can have. In the case of wine tasting, learning concepts corresponding with flavors recognizable as ‘dry’, ‘heavy’ or ‘bright’, enables one to be aware of particular types of qualitative complexities. Similarly, learning to distinguish between different kinds of instruments influences one’s phenomenal experience of musical tones. A chord suddenly sounds different when we learn that it is produced by an organ rather than by a violin. Rosenthal hypothesizes that higher-order thoughts could employ concepts to this end. In ‘Explaining Consciousness’, he writes,

> 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 a sound of a woodwind, what it would be like for one to have that sensation would then be correspondingly more generic (2002, p.413-414).

A third facet of subjective experience that the HOT theory is proposed to help explain is the commonplace phenomenon of self-deception. Self-deception is an event in which one is mistaken about one’s own occurrent psychological state. When we are self-deceived, our experience of ourselves is not consistent with what is really going on psychologically on a more fundamental level. Rosenthal refers to two examples of self-deception. The first is imagined dental pain,

> 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 that the patient experiences fear along with vibration from the drill and consciously reacts as though in pain (2002, p.415).

Repression is another type of psychologically motivated error about the self that Rosenthal refers to. In cases of repression, one unconscious desire is often redescribed and experienced in consciousness as an entirely different desire (Rosenthal 1986, 2000a,
Freud’s patient, Lucy R. manifests a famous example of this phenomenon (Freud 1895). Lucy was an English governess in Vienna who took care of the house and children of a rich widowed man. She suffered from a loss of the ability to smell, and importantly, from olfactory hallucinations centering on the smell of burnt pudding. Freud discovered that she was secretly in love with the man of the house and that she had often made pudding for the children. Instead of experiencing her feelings of attraction to the man, she repressed them and they arose in consciousness as olfactory hallucinations. When Freud pressed her to acknowledge and experience the repressed attraction, the hallucinations of burnt pudding went away.

According to the HOT theory, self-deception is the product of higher-order misrepresentation; in repression, higher-order thoughts represent lower-order states as carrying qualities or information that they do not really carry. When we have higher-order thoughts that misrepresent our lower-order thoughts, we become conscious of ourselves as being in states that we were not actually in. Although we are mistaken about what experiences we were having here, it is consistent with the HOT theory that the misrepresented lower-order psychological state could still play an active role in the cognitive system by influencing thought and behavior below the threshold of conscious awareness. Consequently, this view fits with modern psychological theories of repression.

Although the HOT is powerful in explaining various aspects of conscious experience, some challenging objections have been lodged against the theory. These objections have been influential in arousing skepticism among philosophers concerning this model of consciousness. In an attempt to salvage some of the strengths of the model without inheriting its vulnerabilities, several philosophers have tried to improve on the theory by developing alternative self-representational theories of consciousness. It is my contention that these models do not present a genuine advance over the original higher-order model. But before discussing the way in which self-representational theories are alleged to improve on the original theory, it will help to review some of the standard objections against the original version of HOT.

3. The case against the higher order thought theory

Over the years, some challenging objections have been lodged against the HOT theory. These objections have been influential in arousing skepticism toward the theory.

One standard objection lodged against the HOT theory is known as the generality problem. Block (1995), Goldman (1993) and Dretske (1995) have each articulated well known versions of this criticism. The main issue here is that it does not seem that an intentional relation could be sufficient to make intentional objects conscious. The crux of the criticism is often illustrated by considering an analogy; if the internal states of stones do not become conscious when we are conscious of them, and states of our own body such as our liver do not become conscious when we have thoughts about them, then why should mental states become conscious when we are aware of them? The argument from
analogy suggests that, unless the case of consciousness is radically different from other forms of transitive awareness, it is hard to see why representing a target mental state should have this sort of effect.iv.

Another objection often lodged against the HOT theory states that the possibility of higher-order thoughts misrepresenting their targets renders the HOT theory logically incoherent (Neander 2002, Levine 2001). Suppose that a red lower-order qualitative state is rendered conscious by a higher-order thought that characterizes that state as being blue. Or consider self-deception, where the self-deceiver misrepresents his fear as pain, resulting in his fear feeling to him as though it were pain. One philosopher who has objected to the HOT theory on this ground, Karen Neander (2002), notes that if this account is supposed to be an explanation of conscious ‘qualia’ as traditionally understood, it is deeply puzzling. The theory seems to predict that when we misrepresent a lower-order red qualitative state as blue, what it is like to have a red qualitative experience will have the character of being bluish.

A related problem for the HOT theory concerns the issue of confabulation. Confabulation is, like other types of misrepresentation, a kind of error that results from a breakdown in the mechanism that facilitates the higher-order representation of information. As we have seen, higher-order thoughts can misrepresent their targets. But insofar as a higher-order thought is a type of thought, it seems that in addition to misrepresenting its object, it could occur in the absence of an instantiated intentional object. After all, such confabulations occur in the context of many other types of thoughts. For example, perceivers may hallucinate and dreamers may imagine non-existent fictional entities such as unicorns or elves.

It is natural to wonder what a confabulatory higher-order thought would be like. Critics hypothesize that according to the HOT theory, confabulation would result in no conscious experience at all. Alex Byrne criticizes the theory when he writes, “I could have a higher-order thought that I am having a visual experience of a tree, but without having a visual experience of any kind.” (1997, p.122). Along a similar line, Georges Rey (2000) has noted that in such cases, there doesn’t appear to be any state that we could properly say is intransitively conscious. And Katalin Balog notes that confabulation entails the ridiculous possibility of a HOT zombie, “a creature whose inner life, from the point of view of her stream of consciousness, is identical to mine, except she doesn’t have any sensory states, she never has pains, might not even have ever had any” (2000, p. 218). If confabulations involve higher-order thoughts without any contents or qualities of experience, then higher-order thoughts cannot be sufficient for consciousness.

I have reviewed several standard objections that are often employed against the HOT theory: the generality problem, an inverse problem raised by confabulation, and the problems raised by the possibility of other types of misrepresentation. Assuming that these objections get things right, if the HOT theory is to be at all plausible, then something will have to be changed. Taking this realization as a point of departure, several philosophers have made efforts to modify the original higher-order model in a way that
will retain the basic idea of the metacognitive model of consciousness but at the same
time yield a theory that is less vulnerable to objections.

4. Alternative self-representational theories of consciousness

To accommodate the objections to HOT noted in the previous section, some philosophers
have developed alternative self-representational theories of consciousness by modifying
what they take to be a key aspect of the original HOT theory. Rather than taking the
representing state and the conscious state to be independent of one another as the HOT
allegedly does, Gennaro, Kriegel and Van Gulick’s self-representational theories make
the representing state and represented state two parts of a larger composite state.

First, Gennaro argues for a metacognitive theory of consciousness that he calls the Wide
Intrinsicality View, or WIV. According to WIV, a conscious state is a conglomerate
consisting of a lower order represented state and a metacognitive representation of that
state. He agrees that it seems that something over and above our physically determined
sensory abilities influences how phenomenal qualities are experienced. Gennaro thinks
that the HOT theory is right to explain the characterization of phenomenal experiences
that we saw in the examples of wine tasting and the woodland scene, by concepts
deployed in metacognitive thoughts about those qualities. He writes,

The nature of a conscious state is, in part a function of the concepts which figure
into the MET (metacognitive state). The sophisticated wine drinker has different
qualitative states than the inexperienced wine drinker. The metapsychological
‘thought awareness’ involved in having a MET actually changes the nature of the
conscious state. (1996, p.29)

But Gennaro maintains that WIV is better positioned than HOT to account for the
metacognitive state’s role in conceptually characterizing phenomenal experience. He
argues that, whereas the HOT theory cannot explain how a distinct state could come to
cause a change in the intrinsic qualitative properties of the conscious state that it
represents, the fact that WIV unites the metacognitive state and the target into a larger
sum state, enables the MET to characterize phenomenal experience.

The MET is not a distinct state directed at another state. On the [HOT] theory, the
state which is rendered conscious is viewed as an independent existent and…in
treating the mental state and its MET as distinct, the fact that the MET contributes
essentially to the very qualitative character of the conscious state is lost (1996,
p.30)

Next, the HOT theory’s alleged vulnerability to the generality problem, the problems of
confabulation and misrepresentation, among other objections, motivates U. Kriegel to
develop the Same-Order Monitoring theory (SOMT). SOMT is centered on the following
principle,
SOMT: A mental state M of a subject S is conscious iff (i) M* is a (proper) part of M, (ii) M^ is a (proper) part of M, (iii) M* is a representation of M^, and (iv) M is a complex of M* and M^ (Kriegel 2005).

According to SOMT, the represented and representing components of a conscious state are integrated with one another through a special process of information integration. Consciousness is a property that a mental state has in virtue of the relations that obtain among the integrated parts. In making consciousness a property that states have by virtue of the relations between their own parts, rather than by virtue of relations to other states, SOMT makes consciousness to be an intrinsic property of mental states rather than a relational property. Kriegel goes on to argue that SOMT is substantively different from, and an improvement over, both WIV and HOT insofar as the integration of information results in the production of a complex state rather than a sum.

Complexes differ from sums in two important ways. First, a complex is a state whose properties add up to more than the properties contributed individually by each of its parts; that is to say that complexes have emergent properties. At least one of SOMT’s emergent properties is a causal property. Kriegel states that unlike WIV, SOMT does not make consciousness causally inert,

This can be seen clearly by noting that if the monitored and monitoring states are unified through a psychologically real process, that process would presumably make a difference to the causal power of the complex of the two, something that would not happen if the monitored and monitoring states are simply summed up (2005, p.151).

Second, insofar as a complex is an emergent state with emergent properties, a conscious state’s parts, according to SOMT, are not logically independent of one another. And if the represented state and representing state are not logically independent of one another, consciousness cannot occur where one state occurs in the absence of the other. This is to say that SOMT rules out the possibility of confabulation. Nor is SOMT vulnerable to the generality objection since the only states capable of being conscious will be states that are capable of the proper sort of integration with representing states. Mental states can be integrated in the right way with mental states but probably not with stones. Finally, the concerns that arose for HOT surrounding the event of misrepresentation do not seem to arise for SOMT. SOMT holds that the misrepresenting state is constitutively linked with the conscious state, thereby enabling that information to have a claim in the constitution of phenomenal character.

A third alternative self-representational view is Van Gulick’s, Higher-Order Global States theory (HOGS). According to HOGS, there becomes something that it is like to be in a particular mental state m when m is incorporated into a higher-order global state. Van Gulick claims several other theories as inspiration for HOGS. First, HOGS resembles Dennett’s cerebral celebrity model according to which conscious states are just states whose contents have more influence in the cognitive domain. States that are incorporated into the globally integrated complex will naturally bear more connections than states that
aren’t, adding to their potential to influence those states. Second, HOGS is inspired by Hill’s notion that introspection is an active process that can produce change the nature of the introspected state in various ways, such as by making it more intense. Van Gulick writes, “Integration proposed by HOGS is a dynamic reciprocal process in which the various components of a global state both amplify and modify each other in content sensitive ways.” (2005, p.79)

Finally, the HOGS model is motivated by the fact that the substrates of conscious experiences are, “globally distributed patterns involving many different cortical and subcortical regions that are simultaneously active and bound together in some way”, rather than the more common view that there are unique and isolatable neural correlates of consciousness. The view that there are globally distributed neural correlates fits very well with the assumptions of HOGS. According to HOGS, conscious experience is realized by whatever brain areas realize the mental states that happen to become unified with the higher-order global state. So, according to HOGS, not only is there not one specific area of the brain correlated with conscious experiences, but the same areas that realize non conscious states come to be the substrates of conscious states when the state that they realize becomes integrated with conscious states. This feature of HOGS also makes it invulnerable to the generality problem since it seems only reasonable that just mentally tokened information could be recruited into a higher-order global state.

Finally, like HOT, HOGS has the virtue of providing a way to account for the fact that conscious experiences are accompanied by the sense that those experiences belong to us,

Meta-intentionality derives not from the distinct state but from implicit self perspectuality built into the intentional structure of conscious experience itself…presentness of objects always seems a matter of presence to self and from the self perspective. It follows that understanding the reality of experienced objects requires at least implicitly understanding the nature of the self to which they are present. It is in this respect that conscious experience embodies a significant degree of reflexive meta-intentionality in its very organization and structure. (2005, p.85)

Each of these models appears to constitute an interesting metacognitive alternative to the HOT theory. I plan to show that not one of them presents a genuine advance over the original HOT model. First, there is reason not to hold HOT to be genuinely vulnerable to the objections that are commonly lodged against the theory. Even so, however, it is fair to wonder still whether any of these alternative models might not have some other advantage over the higher-order version. But it turns out that each of the alternatives mentioned is vulnerable to its own set of objections. One can take this as support in favor of the traditional higher-order model over the alternatives discussed.

5. Two models of the higher-order theory
It is not clear that we should consider the HOT theory to be vulnerable to the standard objections made against it. The assumption that the theory is vulnerable to those objections comes from holding one interpretation of the higher-order model of consciousness. But there is another way to interpret the theory that seems to be better supported by the arguments for HOT.

Recall that the central idea of the HOT theory is that a mental state become conscious when the subject it belongs to has a roughly contemporaneous thought to the effect that she is in that very mental state. But this very general statement can be interpreted in two different ways based on whether the term ‘mental state’ is thought of as referring to mental-state-types or to mental-state-tokens (Matey 2006). ‘Mental state’ refers to a mental-state-type if it picks out or refers to a specific but abstract property, proposition or universal. On the other hand, the term ‘mental state’ refers to a mental-state-token qua token if it is used to pick out a particular exemplification of a type instantiated at a time and place such as the pain Hercules had due to his kidney stone on Wednesday. Whether one interprets ‘mental state’ here as referring to mental-state-types or to mental-state-tokens has important implications for the specific way in which the theory would be modeled. Consider the following formulation of the main tenet of HOT,

\[ \text{A mental state} (1) \text{ is conscious...just in case one has a thought} (2) \text{ to the effect that one is in that mental state} (3). \]

The way that we interpret (1) should determine the way that we interpret (3) since the use of ‘that very’ strongly suggests that (1) and (3) are one and the same entity. The main problem as I see it is to determine the intended meaning of these first and the third instances of ‘mental state’.

By interpreting ‘mental state’ there to refer specifically to tokens, the transitivity principle reads as follows,

\[ \text{Token TP: A mental-state-token becomes conscious...just in case one has a thought to the effect that one is in that very mental-state-token.} \]

This version of the transitivity principle states that the lower-order target state token becomes conscious. It seems fair to assume that when a theory refers to a particular state token becoming ‘conscious’, then it is the informational properties or qualities instantiated in that token that we are to understand as constitutive of phenomenal or qualitative character. This reading resembles the intuitively appealing understanding of consciousness as where a state acquires the property of phenomenality, but here we need not commit to the view that consciousness consists in acquiring irreducible phenomenal properties. If a set of qualities is constitutive of phenomenal character, then those qualities determine the phenomenal character. It should be noted that in saying that character is determined by qualities or information present in a particular mental state, we do not merely mean that those qualities determine the fact that there is phenomenal character. Rather, those qualities or informational properties determine precisely what the
phenomenal character in that instance will be like. So Token TP yields the following model of consciousness,

M1: Phenomenal consciousness consists in higher-order mental-state-tokens representing lower-order mental-state-tokens and the phenomenal character of the experience is determined by qualities or informational properties as instantiated in the lower-order token.

Alternatively, the term ‘mental state’ in the transitivity principle can be interpreted as referring to a mental-state-type,

Type TP: A mental-state-type becomes conscious...just in case one has a thought to the effect that one is entertaining that very mental-state-type.

Prima facie, the claim that mental-state-types can become conscious may sound a bit strange. This superficial problem disappears, however, when we consider more familiar but similar locutions. For example, it would not seem strange to ask someone whether or not they have ever experienced kidney stone pain. Here we are not asking about whether they have had a particular pain token, but rather about whether they have ever had a token of a particular type.

Type TP says that a mental-state-type which may have already been exemplified in a non-conscious lower order state token, becomes conscious, which is to say that there becomes something that it is like to entertain it, when one has a higher-order representation indicating that one is entertaining that very mental-state-type. But Type TP does not require that the mental-state-type that comes to be phenomenally conscious be instantiated twice. Call the higher-order representing state M* and the lower order unconscious token state M. According to Type TP, M* may misrepresent M, it may characterize M in some way that yields a relationship less than identity, or it may even occur in the absence of M. This is to say that there need not be both an M* and an M that fall under T in order for T to be conscious, but only some token M* which represents that one is in mental state T. Type TP, therefore, yields the following model,

M2: Phenomenal consciousness consists in a higher-order mental state representing that we are in a particular mental state, and phenomenal character is determined by qualities or informational properties instantiated in the higher-order token.

To summarize, the way that the term ‘mental state’ is interpreted in the context of the transitivity principle influences the way that the theory is modeled. ‘Mental state’ could be interpreted as referring to mental-state-types or mental-state-tokens. Interpreting ‘mental state’ as referring to mental-state-tokens yields a model that locates the phenomenal character determining powers in the lower-order state token. Interpreting ‘mental state’ as referring to mental-state-types yields a model that locates the phenomenal character determining powers in the higher-order representing state token. In light of this distinction, it is interesting to investigate two questions.
First, which model, M1 or M2, is best motivated by the arguments offered in support of the theory? The second question that comes to mind is whether either of these models is more vulnerable to the standard objections typically lodged against the HOT theory. I want to propose that M2 is the model best motivated by the original arguments offered in support of HOT, but M1 and not M2, is vulnerable to the standard objections that have been lodged against the HOT theory. The natural conclusion to draw here is that the version of the HOT theory best supported by the literature on HOT is not vulnerable to the objections that have caused so much skepticism toward HOT. This can be summed up in the following argument,

1. M1 but not M2 is vulnerable to the standard objections.
2. M2 is the version of HOT best supported by the arguments for HOT.
3. We should select the strongest interpretation of the HOT theory.

C: We should interpret HOT as M2.

If HOT holds up in the face of standard objections, we have that much more reason to consider it to be a good contender for a theory of consciousness, and that much less motivation for adopting one of the alternative self-representational theories in its place.

6. The higher-order-thought theory and the standard objections

Only model M1 is vulnerable to many of the standard objections made against the HOT theory. But M2 is the version of HOT best supported by the arguments for HOT. Therefore, we should not take the HOT theory to be vulnerable to the standard objections.

It will help to spell out the reasons that we have for accepting this argument’s premises.

(1) The standard objections to the HOT theory target M1 but not M2.

Three of the objections most commonly lodged against the HOT theory are: the generality problem, the problem of misrepresentation, and the possibility of higher-order confabulation. These objections have been influential in arousing skepticism directed at the HOT theory.

Consider the generality problem. Block (1995), Goldman (1993), and Dretske (1995) allege that the HOT theory makes consciousness consist in the obtaining of an intentional relation. But intentional relations are not of the sort to typically make an intentional object conscious. This becomes evident when we consider analogous cases. The internal state of a stone does not become conscious when we are conscious of it, nor do states of the liver or veins. What should make mental states different?

The absurdity derived from the analogous cases rests on the assumption that the target’s intrinsic qualities or properties become those constitutive of phenomenal character. This is a way of construing M1; if the target state’s qualities are constitutive of phenomenal
character, then they also determine phenomenal character. On the other hand, if the higher-order state were to determine phenomenal character, as M2 would have it, then it is hard to see why we should be concerned about the generality problem. We would not have to worry about how states could bring about a change to the intrinsic qualitative properties of their intentional targets.

Next, consider the problem of misrepresentation. Neander noted that, if the mechanism responsible for producing HOTs were to break down, according to the HOT theory, a red lower-order qualitative state could be rendered conscious by a higher-order thought characterizing it as blue. This has the deeply puzzling result that what it is like to have a red experience will have the phenomenal character of bluishness. But this paradoxical result only arises if misrepresentation is understood within the context of M1 where phenomenal character is supposed to be determined by the target red state. For M2, there is no deeply puzzling problem surrounding the phenomenology of misrepresentation. If the character of our experience is determined by the higher-order representing token, then in the case that we misrepresent a pain as fear, we would feel as though we are afraid. The fear type is conscious in virtue of a higher-order token. So M2 is consistent with folk-psychological descriptions of self-deception.

Third, consider the problem of possible confabulatory higher-order thoughts. If higher-order thoughts are the result of a cognitive mechanism and mechanisms can break down, then it seems possible that one’s cognitive system could produce a higher-order thought even in the absence of a lower-order-target. The problem with such confabulatory higher-order thoughts, according to some, is that without a represented target state there wouldn’t be anything that it is like to have a the higher-order thought. Alex Byrne writes, “I could have a higher-order thought that I am having a visual experience of a tree, but without having a visual experience of any kind” (1997, p. 121-122). If higher-order thoughts can occur without there being something that it is like to have them, then consciousness qua what it is likeness, cannot consist in mental states being represented by a suitable higher-order-thought. Notice how the force of this objection comes from the assumption that it is the role of target states to determine phenomenal character. Confabulation is not a problem for M2, since the higher-order token is sufficient to determine phenomenal character even when no lower-order token exists.

M2 appears to be superior to M1 insofar as it is not vulnerable to common objections lodged against the HOT theory whereas M1 is vulnerable to these objections. If it can be shown that M2 is better supported by the HOT literature, then we will have removed a major ground for skepticism toward the HOT theory.

(2) M2 is the version of HOT best supported by the arguments for HOT. M2 fits better than M1 with the motivating examples that the HOT theory was put forward to explain. In accord with the principle of charity, we should assume the model of the theory that would best explain the phenomena that the theory purportedly explains.
For example, the HOT theory offers to explain the way that it seems that concepts circumscribe the experience of the qualitative features that we perceive. David Rosenthal cites examples of the influence of concepts on the taste of wine, the experience of music, and also the difference they potentially make for how we experience background scenes as in the case of the woodland. In these examples, it seems that something over and above our sensory abilities influences the way that conscious qualities are experienced. The HOT theory explains this by making it the result of top-down influences on our perception of those qualities. The problem is to explain exactly how this top-down effect could occur.

Top-down influences indicate M2 over M1. The HOT theory makes consciousness consist in the obtaining of a special kind of intentional relation. Moreover, in cases such as the wine tasting example, the higher-order state influences phenomenal character by influencing the way that non-conceptual information presented is experienced. The view that the qualities present in the lower-order state token determine phenomenal character cannot explain how the higher-order state could come to influence phenomenal character without making the relation between the higher-order thought and its target consist in more than a mere intentional relation. This problem is highlighted by R. Gennaro (1996) who charges that the HOT theory will have difficulty explaining how a distinct higher-order state could come to cause a change in the intrinsic qualitative properties of the conscious state that it represents. This is only a problem if we assume that the qualities as they are tokened in the target state determine phenomenal character.

On the other hand, the view that higher-order states determine phenomenal character makes better sense of the examples from wine tasting and the woodland scene. Invoking the principle of charity, where a theory supports two models we should assume the version of the theory that makes sense of the motivating evidence.

Second, the HOT theory is proposed to explain the cognitive mechanisms involved in common types of self-deception. The self-deceived individual experiences herself as being in states that she had not actually been in. In repression, one state is experienced in the place of another state that there is some motivation to not experience. For example, the dental patient experiences herself as being in pain rather than as being afraid, and Lucy R. experiences the smell of burnt pudding rather than her scandalous desire. M1 and M2 do not provide equally plausible models for self-deception.

M1 can only make sense of self-deception experiences by making it to be the case that higher-order misrepresentations cause target states to undergo a change in the qualitative properties that they are comprised of. In misrepresentation, phenomenal experience is consistent with the way that the higher-order state characterizes the target, and M1 has it that target states determine phenomenal character. Such a change amounts to altering the state’s intrinsic nature. Again, it is unlikely that a mere intentional relation between two independent entities could be sufficient to cause a change in the represented state’s intrinsic properties.
M2, on the other hand, offers a coherent explanation of self-deception and other types of misrepresentation. According to the folk understanding of self-deception, we experience ourselves as being in states that we were not actually in. Self-deception makes sense in the context of M2, where phenomenal character is determined by the way that the representing state characterizes our experience; in misrepresentation whatever state we were in is mischaracterized. Again, invoking the principle of charity, we should assume the model that makes the best sense of the motivating evidence.

Third, consider confabulation. Rosenthal claims that higher-order thoughts might even arise when the state that they purport to represent does not exist. Moreover, he suggests that in the context of such confabulations, phenomenal character is determined by the higher-order representation. He writes,

Confabulated conscious states are states we are conscious of ourselves as being in even though the states do not actually occur. We are, in this way, actually conscious of states we aren’t in, but subjectively seem to us to belong to our stream of consciousness…being conscious of a state does not imply the state exists (2000a, p.211).

Rosenthal claims that confabulatory higher-order thoughts make ‘notional’ states intransitively conscious (2000). A notional state is a mental state type that is not tokened. To assume M1 would be to leave Rosenthal in the awkward position of having to account for how there could be something that it is like to be in an untokened mental state. A more charitable assumption would be that phenomenal character is determined by higher-order tokens. M2 explains confabulation as a mental event in which representations of notional target states, qua types, come to be tokened in higher-order thoughts, even if not in lower-order targets.

Since, in cases of confabulation, phenomenal character goes with the higher-order state, it seems most parsimonious to assume that phenomenal character goes with higher-order state in other cases as well. The alternative would be a disjunctive view where M1 obtains in typical cases, but where M2 obtains in the case of confabulation. The principle of charity demands that we look for parsimony.

It has been shown that, if we invoke the principle of charity, we should assume the M2 model of the HOT theory rather than M1. M2 better explains the phenomena that the HOT theory was proposed to explain. But what is more, M2 provides a potentially illuminating explanation of additional cognitive phenomena such as tip-of-the-tongue, whereas M1 does not. The fact that M2 can explain tip-of-the-tongue adds to the model’s strengths. Again, invoking the principle of charity, we should assume the model that has greater overall strengths.

Tip-of-the-tongue is an experience in which we have mental states that inform us about the fact that we are in or have a particular information carrying state, yet we do not have conscious access to that state. From the subject’s point of view, a tip-of-the-tongue experience is of the sort where it is as if some relevant information such as a person’s
name is cognitively present, but the subject is unable to access the phonological structure of the cognitive item. Rosenthal describes,

When I have George Orwell’s 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 that state in respect of that information (2000, p.205).

The HOT theory presumes a functional account of information whereby the informational content of a mental state is specified by the state’s causal relations or its functional role. A ‘pain’ state is not merely individuated in terms of how it feels, but also in terms of the relation that it bears to inputs, outputs and other mental states such as beliefs and desires. For a state to carry information about Hemmingway’s name consists in the state playing the right role and having the right relations to other states. Part of playing the right role is for there to be particular questions that the state provides the answer to.

M2 provides a potentially illuminating explanation of the tip-of-the-tongue phenomenon. According to M2, the tip-of-the-tongue state, “I am in a state that answers to the question ‘who wrote Animal Farm’?” would be an instance in which a higher-order state partially represents the informational content of its intended target. I represent the state in virtue of some role it plays, but not in virtue of its exact phonological structure. According to this view, regular consciousness conferring higher-order thoughts and tip-of-the-tongue states differ only with respect to which particular informational properties of target states they represent. The fact that it is precisely the specific piece of information that we are seeking in tip-of-the-tongue, a description or proper name, that is not conscious makes it seem to be that the difference between higher-order thoughts and tip-of-the-tongue states is a difference in kind rather than a difference in degree.

M1, on the other hand, leaves us to wonder just what makes for the experiential difference between regular consciousness conferring higher-order thoughts and tip-of-the-tongue states. M1 is the view that when metacognitive states represent targets, phenomenal character is wholly determined by the lower-order state. M1 predicts that lower-order tokens will determine phenomenal character regardless of the way that they are represented, even when the mechanism responsible for generating higher-order thoughts breaks down. If the lower-order state token determines phenomenal character regardless of the way the token is represented, then how do we explain why there should be tip-of-the-tongue experiences?

I have discussed two models of the HOT theory that result from taking the term ‘mental state’ in the transitivity principle as referring either to mental-state-types or to mental-state-tokens. These two models were discussed in respect of how they handle several standard objections to the HOT theory and also in respect of how well they meet up with the theory’s main aims and motivations. It was shown that M1, but not M2, is vulnerable to the standard objections lodged against the HOT theory, but that M2 and not M1 fits with the HOT theory’s supporting arguments by explaining key aspects of subjective experience. Applying the principle of charity, we should read the HOT theory as M2. So
not only does HOT appear to be well motivated, but it also appears to be without significant objections as far as we can tell.

7. Weaknesses of alternative self-representational theories

It is interesting to consider how the original higher-order model compares against alternative self-representational models of consciousness in light of section 6’s conclusions about HOT. Authors of alternative self-representational theories have professed the aim of salvaging what is meritorious from the original higher-order model while avoiding the objections that the HOT theory appeared to be vulnerable to by making the representing state and the state that determines phenomenal character, constitutively linked. We have found that there are reasons to read the HOT theory as not actually vulnerable to the objections often lodged against it since at least one viable model already links the representing state with the state that determines phenomenal character.

But even if these self-representational theories do not improve on the higher-order model in the way that they have been proposed to, this does not mean that they are not superior to the HOT theory due to some other feature. In this final section, I want to show that this is in fact not the case. Each of the alternative theories reviewed is vulnerable to its own set of objections, making each of them unlikely to constitute a genuine improvement on the original higher-order model.

Gennaro proposed the Wide Intrinsicality View. He argued that, in making the target state and the metacognitive state parts of a larger sum, his theory does not run into the problem of having to account for how an independent higher-order representing state could have an effect on the qualitative nature of the state that it represents. Problematically, however, it isn’t clear that Gennaro’s WIV is substantively different from HOT. According to WIV, the higher-order representation and its target are united into a composite state in which the parts are summed-up. In a whole comprised of summed-up parts, the parts remain logically independent of one another.

Whenever we have a single entity comprised of logically independent parts, we must presuppose some theory of mental state individuation to explain why the independent parts should be regarded as constituting a single ontologically novel entity. One way to individuate entities appeals to causal powers; something is considered ontologically novel if and only if it has unique causal properties. Unfortunately, in the case of WIV, we are not provided with any reason to think that a composite conscious state would have causal properties over and above those contributed independently by each of the state’s parts.

In Gennaro’s defense, it might be proposed that we would do well to appeal to a pragmatic theory of state individuation in which states are individuated to the extent that such an ontological description yields more explanatory power, less inconsistency or fewer errors. But in light of what we have discovered about HOT qua M2, the description of consciousness offered by WIV does not seem to constitute an improvement over HOT.
on any of these grounds. Both HOT and WIV seem equally well motivated and neither are vulnerable to the standard objections. In summary, there is no reason to think that WIV differs from the higher-order model in any way that makes a difference.

The second alternative self-representational theory proposed in the place of the HOT theory was R. Van Gulick’s, Higher-Order Global states model (HOGS). HOGS makes consciousness consist in the relevant state becoming incorporated into a higher-order global state. Since only mentally tokened information would be capable of being recruited into a global mental state, the generality problem does not arise for HOGS. On the other hand, as we have seen, the generality problem does not arise for the HOT theory either.

Moreover, even if HOGS is not vulnerable to the generality problem, it is still an unlikely contender for a reductive theory of consciousness. First, HOGS is not offered as a reductive explanation of consciousness. Van Gulick himself sees HOGS as more de facto psychological description of events surrounding phenomenal consciousness than reductive analysis,

HOGS focuses on the thick and densely integrated structure of phenomenal intentionality and the fundamentally dual aspect of experience both of self and world. It thus gives us opportunities to try to understand how such a structure of phenomenally intentional content might be realized by a globally integrated state that seems to provide its physical substrates…but I certainly do not claim that it does the job by itself or that it thereby solves the “hard problem”. (2005, p.90)

We might wonder if Van Gulick is too modest in his appraisal of the scope of his theory. It is worthwhile to consider whether, independent of Van Gulick’s own assessment, HOGS may be a viable reductive account of consciousness nonetheless. Since HOGS shares structural similarities with other reductive models, such as Dennett’s cerebral celebrity model, it may be worthwhile to look at Dennett’s model. If Dennett’s model is successful, perhaps HOGS is similar enough structurally to inherit that success.

The cerebral celebrity model equates consciousness with connections to, and a high degree of influence on, other cognitive states and behavior. One intuitive measure of consciousness that Dennett endorses is availability for report. Conscious states are states that we can make judgments concerning. As such, they must be conceptually encoded. Insofar as perceptual experiences can be conscious, this will apply to perceptual experiences as much as to doxastic states. The problem is that when we introspect on our conscious experiences, especially in the case of perception, we find that we are often able to discriminate between qualities that we do not have words or concepts to individuate. Much of experience is what philosophers call ‘belief independent’ (Peacocke 2001, Crane 1992, Byrne 1996, Heck 2000). In short, experience is quite often much richer than the judgments we are capable of making indicate.

Dennett’s response to this objection is unpersuasive. According to Dennett, we do not have a reliable way to access experiences that cannot be verbally reported. The only basis
that people have for claiming that they have non-conceptual qualia is introspection and Den-nett argues that introspection is too unreliable to serve as a source of self-knowledge about experience. Therefore, we must leave anything that cannot be reported out of our account of consciousness. The example of the zombie is used to illustrate this point. Most who think that phenomenal experience outstrips intentional content also hold that insofar as phenomenality is independent of intentionality, it seems intuitive that there could be something just like us in physical and functional respects, but lack our richness in phenomenal character. But if such zombies are really possible, how could I ever know that I myself am not a zombie? How could I determine that my introspective judgments about phenomenal character are true? If my zombie twin is genuinely functionally identical to me, she will be disposed to make the same judgments about her experience as I make about mine. Dennett thinks that it would follow from this that I could not be justified in the judgments that I make about my own qualia.

Problematically however, Dennett assumes that the view that experience outstrips what we can judge goes hand in hand with intuitions about the possibility of zombie-twins. But it need not. If my zombie twin has a belief that she is having a certain experience but she has no experiences from which that belief is deduced, then there is at least one way in which she is not like me- epistemically (Levine 1994). My judgments are true and they are based on experiential evidence. The zombie’s judgments are false and unjustified.

Moreover, it is questionable whether Dennett’s analysis is even an analysis of consciousness in the phenomenal what it is like sense at all (Block 1995). Block takes ‘consciousness’ to be a mongrel concept used to pick out different properties on different occasions of use. Most commonly, ‘consciousness’ picks out either a state’s property of being phenomenally conscious, or a state’s property of being accessible for use in reasoning and controlling behavior. Block notes that Dennett’s cerebral celebrity model is more obviously an analysis of what constitutes a state’s being access conscious rather than an analysis of what constitutes a state’s being phenomenally conscious (1994). The following quote by Dennett seems to corroborate this,

Consciousness is cerebral celebrity, nothing more, nothing less. Those contents are conscious that persevere, that monopolize resources long enough to achieve certain typical and symptomatic effects on memory, on the control of behavior and so forth (1993 p.929).

In summary, construing HOGS in the spirit of Dennett’s cerebral celebrity model does not make it a more convincing alternative account of consciousness since the cerebral celebrity model is not obviously a model about the phenomenon of phenomenal consciousness.

The final self-representational theory to be considered is Kriegel’s Same-Order Monitoring theory of consciousness. According to SOMT, a mental state becomes conscious when it represents itself. Kriegel’s theory is similar to WIV in that it holds that the represented and representing components are constituents of a single conscious state. Unlike WIV, however, SOMT takes it that these two constituents are not merely
summed-up, but rather integrated into a larger complex state. Thus, SOMT, according to Kriegel, is substantively different from WIV and HOT. SOMT also differs from WIV insofar as it makes consciousness to be a matter of the relations that obtain between parts of a state. This enables consciousness to be an intrinsic property of mental states rather than merely an extrinsic matter of relatedness to another state.

Kriegel claims that on account of these unique features, SOMT is immune to several objections often lodged against the HOT theory. One general objection that is sometimes raised against the HOT theory concerns the fact that HOT seems to be incompatible with the intuition that conscious mental states are causally efficacious in virtue of their phenomenal characters. There is a common intuition that the phenomenal aspect of experiences is essential to the causal roles of conscious states. When I put my hand in the flame and then withdraw it quickly, it seems that my awareness of being in pain is part of the cause of my withdrawal of my hand. Dretske (1994) has argued that, according to the HOT theory, the particular way that things feel for me right now cannot have any bearing on my future thoughts and feelings; HOT makes consciousness is causally inert.

In more detail, the objection goes as follows. A causal power is a property that something has in virtue of its intrinsic properties. Since the HOT theory construes consciousness as a relational property, consciousness cannot contribute to a mental state’s causal powers. In other words, since being transitively conscious of a state makes no difference to the state’s non-relational properties, being conscious will make no difference to the state’s causal properties or function.

There are several ways to defend the HOT theory against this objection. First, it could be argued that consciousness really is directly relevant to a state’s causal role. It is therefore no surprise that we would have the intuition that it is. For example, if the HOT theory were seen as bearing the relevant similarities to Baars’ Global Workspace theory (1997) or Dennett’s Cerebral Celebrity model of consciousness (1991) (there is no obvious incompatibility), then it would be consistent with the HOT theory that conscious states would enjoy broader influence within a subject’s cognitive system.

Or, it might be argued that the HOT theory offers a unique way to solve this problem. For example, when a previously unconscious mental state type becomes conscious, being conscious may result in an intensification of its functional properties insofar as the doubling of the information that takes place when higher-order representations represent a target state whose content falls under the same type, intensifies the affective valence of those represented mental state types (Matey 2006). Information is not emotionally or affectively neutral. Even mental states that appear to be without affective content when analyzed semantically, are frequently accompanied by an emotional valence. A state’s functional associations include its links to our varying wishes, hopes, fears, desires, along with emotions such as joy and sadness and to the behaviors it influences.

Pat may like Sam a little or a lot. Although the informational content of a mental state such as the desire to meet with Sam on Saturday either exists or does not, the emotional valence can vary in intensity. If we could rate a desire in accord with its level of intensity,
we might give Pat’s desire to meet Sam at time $t$ a rating of ‘6’ on a 1-10 dimensional scale. When Pat’s desire becomes conscious via a second state that represents that target desire, Pat will experience a single information type, but have two token states with the same informational content. If the states are identical, then each would have a level of intensity with a rating of ‘6’. Pat now has a desire to meet with Sam with an intensity sum from both tokens of 12. A 12 might influence Pat to do things that a 6 would not.

Moreover, an apparently neutral thought such as that “I must go to the dentist” is associated with other mental states that would have their own emotional valences. So even when a mental state seems to be emotionally neutral, it is likely that a rather complicated set of emotional charges obtain in virtue of the state’s various connections. Moreover, the repetition of “I must go the dentist” in virtue of a higher-order thought entails the repetition of all the connections entailed by the state. Even information with no obvious affective valence may differ in efficaciousness when it is represented by a higher-order thought\textsuperscript{xv}. In summary, there are ways that relational theories of consciousness such as the HOT theory might be capable of accounting for the unique causal powers that we are tempted to attribute to consciousness on the basis of intuitions\textsuperscript{xvi}.

A final alternative defense of the HOT theory against the causal efficacy objection involves disarming the objection. The HOT theory of consciousness assumes that mental states are individuated in terms of their functional roles. Non-conscious ‘pain’ states are defined in terms of their various inputs, associations with other states, and their behavioral outputs. When information is represented in phenomenal character, we should therefore expect that an aspect of the character would include a peripheral awareness of the state’s causal and functional associations. What it feels like to be in pain consists in what it feels like to manifest these various causal relations and behavioral dispositions. A conscious pain is not just an awareness of damage at a specific bodily location, it is also an awareness that I have an inclination to pull my hand away. Perhaps this phenomenon makes the intuition that consciousness confers causality alluring, even if it is ultimately incorrect. The intuition may arise due to the fact that in consciousness, we experience the functional nature of the represented state.

In summary, the fact that Kriegel claims that SOMT is not vulnerable to the causal efficacy problem does not necessarily demonstrate the superiority of SOMT over the HOT theory, since there are several ways to defend HOT against the causal efficacy objection.

Another reason, according to Kriegel, that we should prefer SOMT to HOT is because SOMT does not land us with the problem of having to account for how an intentional relation could result in the right kind of change in the represented state’s qualitative nature. Insofar as SOMT makes a conscious state to be a complex state with emergent properties, a conscious state’s parts, according to SOMT, are not logically independent of one another. It was the independence of the representing state and the conscious state that made both WIV and M1 of HOT vulnerable to this problem.
Unfortunately for SOMT, however, the very factors that enable it to improve on M1 and that distinguish it from WIV, are also the factors that are responsible for the theory’s weaknesses. Two concerns for SOMT stand out. First, SOMT relies on the problematic notion of a logical part. Kriegel admits that the notion of logical parthood is undeveloped, at best. But he claims that there is a prima facae reason for believing in logical parts, as we are perfectly capable of pointing to examples of them.

There are complicated questions surrounding the explication of the notion of logical parthood, questions to which justice cannot be done here. But an example may suffice to illustrate the nature of the logical part-whole relation. When I am glad that the weather is nice, I necessarily also believe that the weather is nice; it is impossible to be glad that the weather is nice without also believing that this is so. But my belief that the weather is nice is not an extra mental act, which occurs in addition to my gladness. In other words, my belief is part of my gladness, in a logical sense of part of. So my believing that the weather is nice is a logical part of my being glad that the weather is nice…that illustration suggests that there is a viable notion of logical parthood that does apply to mental states; it is just that explicating this notion is not easy. (2005, p. 146, 151)

I agree that this is a thorny issue and intend to leave aside the concern about logical parthood in order to focus on what I see to be a more interesting problem that arises for SOMT. That issue concerns whether SOMT provides a genuinely reductive account of consciousness. The motivation for representational theories of consciousness in general has traditionally been to provide a reductive theory of consciousness where consciousness is accounted for in terms of some other cognitive phenomena without appeal to irreducible phenomenal properties. Representationalists contend that ultimately, these reducing cognitive phenomena will prove amenable to physical reduction. If it turns out that SOMT is not compatible with reduction, then regardless of whatever other merits it has, SOMT is not an alternative reductive theory; it is not in the same category of theories as HOT.

Kriegel distinguishes SOMT from both WIV and HOT on the grounds that according to SOMT, the representing component and the represented component are two parts of a larger complex state. A complex, according to Kriegel, is a state that is more than the sum of its parts. It is natural to wonder in what way the self representational state differs from a sum of its parts. Kriegel proposes that at least one way in which it differs is that it has additional properties over and above those contributed individually by each of its component parts. In particular, Kriegel describes conscious states as having novel, emergent causal properties.

Presumably, the novel causal properties that conscious states have in virtue of being conscious will end up having a downward influence on the trajectory of the material processes underlying cognition. Kriegel introduces SOMT’s novel causal properties in a discussion about the theory’s ability to accommodate the intuition that consciousness is relevant to mental causation. But ultimately, the incorporation of the idea that conscious states are complexes rather than sums is a double edged sword for SOMT. While it would
make SOMT substantively different from WIV and M1, in a way that enables it to overcome standard objections to the HOT theory, SOMT relies on the notion of emergence and genuine reduction requires the denial of emergence.

Moreover, if SOMT takes conscious states to be emergent states, then it is not obvious that SOMT is incompatible with property dualism. In summary, if this is what the strongest model of Kriegel’s theory entails—i.e. the model that makes SOMT substantively different from WIV, then Kriegel’s SOMT account should not be taken to be a reductive account of consciousness. It appears that in order for Kriegel to differentiate his theory from WIV and thus from the HOT, he relies on a metaphysics of consciousness that puts SOMT outside of the realm of reductive models of consciousness.

Insofar as the HOT theory is modeled as M2, alternative self-representational theories fail to improve on the theory in the ways that they are proposed to. Moreover, these self-representational alternatives fail to provide superior reductive accounts of phenomenal consciousness insofar as each suffers from its own unique flaws. The WIV model is better seen as a re-description of the higher-order model of consciousness than as an alternative model. HOGS presents an interesting de facto understanding of the dynamic involved in conscious experience, but it is not clearly offered as a reductive solution to the hard problem. And the SOMT model is theoretically interesting but the only incarnation of it that does justice to the notion of genuine emergence puts it outside of the realm of reductive accounts.

**Conclusion**

The traditional HOT theory proves to be the most viable of the metacognitive accounts of consciousness presented. I have suggested that the term ‘mental state’ in the HOT theory’s transitivity principle, might be taken either as referring to mental-state-types or to mental-state-tokens. This results in the possibility of modeling the theory in two different ways. The best way to understand the theory, I argue, is along the lines of M2, a model in which phenomenal character is determined by the qualities or information present in the higher-order representing token, as opposed to M1 where the target determines phenomenal character. The M2 model of the theory is better supported by the initial motivating arguments for HOT as it has more explanatory power than the alternative M1. One consequence of modeling the theory as M2 is that HOT turns out not to be vulnerable to various objections that are often directed at the theory. Moreover, since each of the alternative self-representational theories of consciousness is vulnerable to its own set of objections, none rival the original higher-order model as a reductive theory of consciousness. Many questions remain unanswered and serve as points for future research. For example, it would be helpful to know whether M2’s explanatory power extends to help us to understand cognitive phenomena other than consciousness and tip-of-the-tongue. Moreover, a more extensive study of the causal properties of higher-order thoughts might help us to resolve the matter of the causal efficacy of consciousness.

\[^{xvii}\]
References


---

i It seems also to be the case that we have very rich conscious qualitative experiences, that would seem to outstrip what could be characterized by our conceptual repertoires. If the Higher-Order Thought theory requires that HOTs be conceptual, then accounting for the richness of conscious experience may turn out to pose a problem for the view.

ii To illustrate the phenomenon in question I have appealed to a famous example from Freud’s work. The reader may wonder whether Freud’s work is still relevant and therefore whether such an example illustrates a real phenomenon. Freud’s work is still relevant to contemporary psychoanalytic theory. See Westen 1998. For a discussion of contemporary ideas on repression and suppression of emotion see, Winkielman and Beridge 2004. The traditional Freudian notion of repression where one unconscious emotion or desire is repressed and subsequently redescribed to be experienced in consciousness as something else bears similarity to contemporary psychodynamic analyses of somatization and somatoform disorders. It is thought that dysfunctional processing of affect has a causes subjects to report psychological symptoms of otherwise mysterious origin. See, Lipowski 1987, Bridges and Goldberg 1985, and Taylor Bagby and Parker 1997.

iii For other examples of this objection to the traditional HOT theory, see Gennaro 1996 and Kriegel 2006.

iv The most common response to the generality problem charges that those posing the objection make an implicit and false presumption that consciousness should be construed as an intrinsic property that token mental states acquire. Consequently, target mental states must undergo an internal transformation in order to become conscious. But the assumption that consciousness is an intrinsic property of mental stats is mistaken; the higher-order thought theory construes consciousness as a relational property of mental states.
In “The Same Order Monitoring Theory of Consciousness”, Kriegel distinguishes ten varieties of the approach but I have chosen to focus on the model believed by Kriegel to be the most promising of the ten.

SOMT also avoids the generality problem to the extent that the information located in inanimate objects is not the kind of information that can be integrated in the appropriate way with transitive mental states. Only mentally tokened information can be.

For an example of a neurobiological theory of consciousness which locates the neural correlates of consciousness in global distributed patterns of activation see Kinsbourne 1988, Dehaene and Naccache 2001,

These two ways of understanding the transitivity principle were first distinguished in Matey 2006. The point requires a shift in the most influential ways of interpreting the HOT theory and I hope that the reader familiar with this debate will suspend familiar assumptions about the traditional interpretation and objections in considering the hypothesis put forth in this paper.

The importance of examining the type/token distinction in the context of other metaphysical theories of phenomenal experience and its relation to the material world can be illustrated by appeal to the psychophysical identity theory, which identifies phenomenal states such as pain with their physical realizers. Identity theorists offer the following principle as an explanation for the pervasive and systematic correlations between physical happenings of the brain and what goes on in the mind:

IP: Mental states are identical to brain states.

But identity theories differ with respect to whether ‘mental state’ in the principle above refers primarily to mental state types or mental state tokens and the viability of the identity theory hinges upon this crucial distinction. Smart (1959) offered a type-identity thesis, claiming that mental state types such as ‘pain’ are identical to brain state types such as the firing of C-fibers. But an unfortunate entailment of the type identity thesis is that only creatures with C fibers can feel pain. This seems patently false. The type-identity thesis is therefore considered to be implausible as it does not allow for multiple realizations of mental state kinds. In order to retain the merits of the identity theory while avoiding the problem that accompanies the type-identity thesis, the theory could be imagined as a token identity thesis in which every mental state token is numerically identical to some particular token physical state, usually a brain state token.

Insofar as the HOT theory is to explain the way that mental states which were already instantiated by a subject, albeit non-consciously, come to be experienced in conscious experience, we should expect that when a subject has a conscious experience, the conscious mental state will be exemplified twice, once in the lower order token, and once in the content of the higher-order representation.

One might object that the HOT theory is meant to be a theory about state token consciousness. But Type TP does tell us what constitutes a mental token’s being conscious; whenever we entertain a mental state type, it always occurs by virtue of our instantiating a token of that type. A state token’s being conscious consists in its being a tokening of a certain type of mental state- a state of the form that one has a belief that one is in a particular mental state type- that type is tokened in the higher-order representation as well as allegedly embodied in a lower order token that the higher-order token purports to represent.
In the interest of being thorough, it should be noted that a disjunctive model is also consistent with Type TP. According to this third model (M3), phenomenal character is determined either by represented tokens or by representing tokens. But the disjunctive model is vulnerable to the sorts of problems that plague disjunctive models in general. For example, it will be shown that M2 arguably provides the best explanation of several facets of subjective experience. But if consciousness is best accounted for by M2 in these particular instances, while there is no particular reason to prefer M1 in others, then in the interest of simplification, this should be taken as evidence that M2 obtains in all instances. Finally, a fourth model, M4, in which phenomenal character is determined by neither state is logically compatible with Type TP, but it also should be dismissed, for this model leaves us unable to account for how phenomenal character comes to be the particular way that it is at any given time. The HOT theory in general is motivated as a theory by its ability to provide a coherent account for why phenomenal character is the particular way that it is. In light of these concerns therefore, I will leave M3 and M4 aside and focus instead on comparing the relative merits of M1 and M2.

Van Gulick himself claims that the cerebral celebrity model equates ‘consciousness’ with accessibility by explaining consciousness entirely in terms of intentionality (interpretability from the intentional stance). What is conscious is just whatever I am disposed to make judgments about, and this turns out to be nothing more than what is accessible for use in controlling thought and action.

Whether Dennett’s model withstands the causal efficacy objection is itself controversial. For example, see O’Brien, G. and J. Opie (1997). for why Dennett’s position is also threatened by epiphenomenalism.

I acknowledge that this proposal is incompatible with the intuition that some readers may have that it is coherent to imagine that there could be a being who is physically and functionally identical to themselves but which lacks conscious experiences altogether. Those who have this intuition need not hold the HOT theory at fault based on the causal efficacy objection. They might instead maintain that although zombies are conceivable, this does not necessarily entail that zombies are metaphysically possible. Or, those with zombie intuitions could dismiss the possibility that Higher-Order Thoughts contribute to the affective valence of mental states in the way described and instead dismiss the causal efficacy objection as based on the illusion that consciousness contributes to a state’s causal powers.

This response to the causal efficacy objection is a reproduction of the argument that I offer in my 2006 paper, “Two HOTs To Handle” originally published in Philosophical Psychology, Volume 19.

The author would like to thank several anonymous reviewers for very helpful feedback as well as the audience at the 2008 Society for Philosophy and Psychology meeting and the following individuals: Jayanta Hegde, Daniel Shargel and Justin Fisher.