Synthese (2008) 160:161–181
DOI 10.1007/s11229-006-9106-0

ORIGINAL PAPER

Same old, same old: the same-order representation theory of consciousness and the division of phenomenal labor

Josh Weisberg

Published online: 14 December 2006

Abstract The same-order representation theory of consciousness holds that conscious mental states represent both the world and themselves. This complex representational structure is posited in part to avoid a powerful objection to the more traditional higher-order representation theory of consciousness. The objection contends that the higher-order theory fails to account for the intimate relationship that holds between conscious states and our awareness of them–the theory 'divides the phenomenal labor' in an illicit fashion. This 'failure of intimacy' is exposed by the possibility of misrepresentation by higher-order states. In this paper, I argue that despite appearances, the same-order theory fails to avoid the objection, and thus also has troubles with intimacy.

Keywords Consciousness · Same-order representation · Higher-order representation · Misrepresentation · Qualia

1 Introduction: SORT vs. HORT

The same-order representation theory of consciousness holds that conscious mental states have a complex representational structure—that they represent both the world and themselves.[1] Defenders of the same-order representation theory (SORT) are

[1] This terminology is inspired by Kriegel (2006). Kriegel uses the broader term 'same-order monitoring theory' (SOMT) to cover all views that involve an internal, noncontingent relation between conscious states and subject's awareness of them. My term refers to the subset of those views employing a single, complex *representational* state to explain this relation. Thanks to Uriah Kriegel for helpful remarks clarifying this issue and many other issues throughout the paper.

A version of this paper was presented at the 'Self-Representational Approaches to Consciousness' conference, Center for Consciousness Studies, University of Arizona, March 18th–20th, 2005.

J. Weisberg (✉)
Department of Philosophy, CUNY Graduate Center,
365 Fifth Avenue, New York,
NY 10016, USA
e-mail: jwsleep@aol.com

motivated (at least in part) by a desire to preserve features of the more traditional higher-order representation theory of consciousness (HORT), a view developed in various ways by David Armstrong, David Rosenthal, and William Lycan.[2] However, according to SORT theorists, HORT faces crippling counterarguments. SORT is designed to save key features of the higher-order view, while avoiding its shortcomings.[3]

The key features of HORT worth saving, according to SORT theorists, are HORT's initial characterization of consciousness itself, and the theory's representational approach to explaining consciousness thus characterized. According to HORT theorists, a mental state is conscious when and only when a subject is appropriately *conscious of* being in that state. HORT theorists hold—and SORT theorists generally agree—that if we are in no way conscious of being in a mental state, that state is not intuitively a conscious state. This is taken as a characterization of the main commonsense datum which a theory of consciousness must explain.[4]

In a recent work spelling out his same-order approach, Robert Van Gulick stresses the importance of this characterization of consciousness. He writes,

> It accords well with our common use of the adjective 'conscious' as applied to mental states . . . Our unconscious wants, desires, and memories are hidden from us; they lie 'buried' beneath our awareness, and bringing them to consciousness is a matter of becoming aware of them . . . In so far it is this distinction that it aims to capture, the higher-order view seems pretty much on target (2004, p. 69).

In a similar vein, SORT theorist Uriah Kriegel remarks that,

> To my ear, there is something artificial in calling a mental state conscious when the subject is wholly unaware of its occurrence. Conscious states are not subpersonal states, which we may have, as it were, unawares. Furthermore, it is unlikely there could be anything it is like for a subject to be in a mental state she is unaware of being in (2003b, p. 105).

SORT theorist Rocco Gennaro also embraces the HORT characterization of consciousness, noting that his theory, like HORT, takes

> very seriously the intuitive notion that a conscious mental state M is a state that subject S is (noninferentially) aware that S is in . . . By contrast, one is obviously not (immediately) aware of one's unconscious mental states (2006, p. 222).

This characterization—which, following Rosenthal (2000), I'll label the 'transitivity principle'—provides the theory with the means of distinguishing conscious from nonconscious mental states. Conscious states, as opposed to nonconscious states, are ones

[2] See Armstrong (1968, 1981); Rosenthal (1986, 1997, 2002a); Lycan (1987, 1996, 2001).

[3] SORT theorists of course have other motivations for their view. See Kriegel (2003a); Gennaro (1996); and Van Gulick (2004, 2006).

[4] So called 'first-order representationalists,' such as Dretske (1995) and Tye (1995, 2000), do not accept this characterization of the commonsense data requiring explanation. They hold instead that conscious states make us aware of the world. Furthermore, HORT theorists disagree about the status of this characterization. Lycan (2001) holds that it is stipulative; however, he stresses that the stipulation does correspond to at least one of our ordinary uses of 'consciousness' (otherwise, the stipulation would be uninteresting). Rosenthal (1986, 2000, 2002a, etc.) holds that the characterization is the best way to capture our commonsense notion of mental state consciousness. For the purposes of investigating the differences between HORT and SORT, I will assume, along with the SORT theorists under consideration here, that HORT theorists are correct about the data, and I will assume that it corresponds to at least one way that the folk characterize consciousness.

that we're conscious of being in.[5] The characterization also strongly suggests a 'representational' approach to explaining consciousness—an explanation of consciousness in terms of representations that are about the subject's own mental states.[6] HORT explains consciousness by positing a distinct 'higher-order' representational state in virtue of which the subject is conscious of her conscious states. This higher-order representation represents to the effect that, "I, myself, am in a mental state with such-and-such features." In doing so it makes the subject conscious of that state, and thus the represented state becomes conscious.[7] HORT and SORT both employ representation to explain a subject's awareness of her conscious states; they disagree over whether the representation is distinct from the state it's about, or a part of the state it's about.

HORT offers a relatively straightforward path to a reductive explanation of consciousness, an explanation of conscious states in terms of nonconscious representational elements. According to HORT, consciousness occurs when the right sorts of nonconscious representational processes occur; specifically, when a higher-order representation makes us aware of a nonconscious first-order target state. The higher-order state itself is nonconscious unless targeted by a third-order state.[8] Thus consciousness is explained in terms of nonconscious representations. And since representation is widely seen as amenable to a physicalist explanation, this offers the prospect of what Van Gulick calls "a potentially demystifying account of consciousness" (2004, 69). This is an additional virtue of the HORT worth saving, according to SORT theorists. The SORT theorists I will be considering here all have this kind of demystifying account of consciousness as their general theoretical goal.

However, all three of the SORT theorists mentioned contend that there are serious problems with HORT. I will not review all of their objections here (and I won't rehash the HORT responses),[9] but instead I'll focus on the worry which leads most directly to SORT—one which seems most damaging to HORT and most effectively dealt with by SORT. At a first pass, the objection charges that HORT fails to properly account for the intimate (perhaps constitutive) connection between conscious states and our awareness of them. This lack of intimacy is best brought out by considering the possibility of *misrepresentation* by higher-order states. HORT theorists allow that higher-order misrepresentation is possible. But this exposes a fatal lack of intimacy in the theory, according to the objectors. Avoiding this problem is one of the main motivations for SORT. The objection has been presented in detail by Byrne (1997), Neander (1998), and Levine (2001), and SORT theorists cite it as a major stumbling

[5] Or *aware of* being in—I'll use the two terms interchangeably.

[6] I am using the term 'representational approach' only to cover views that employ a representation directed at one of the subject's own mental states to explain consciousness. My usage thus excludes 'first-order representational' (FOR) views, like Dretske (1995) and Tye (1995, 2000).

[7] There are disagreements between HORT theorists over the nature of this representational state—whether it's thought-like or perception-like—but such differences will not affect the main point here, so I'll ignore them. Likewise with questions about the nature of the reference to the self in the HO content. For a recent survey of this debate see Van Gulick (2001). See also Rosenthal (2004a), Lycan (2004) and Carruthers (2000).

[8] This offers an explanation of introspection, inter alia. For more on this feature of HORT, see Rosenthal (1997, Sect. IV, 2004b). See also Gennaro (2006).

[9] See Kriegel (2003a, 2006); Gennaro (2004, 2005); and Van Gulick (2004, 2006), for a range of objections. See Rosenthal (2002a, 2004a); Lycan (1996, 1998, 2004), for a range of replies. See also Seager (1999), for a good overview of HORT and the debates surrounding it.

block for HORT.[10] Indeed, some SORT theorists have helped to develop and sharpen the worry (Kriege, 2003a). In the next section, I'll lay out the intimacy objection in detail, and in the following sections I'll consider the various SORT responses to the problem. I'll argue that the SORT responses fail to address the worry, and therefore the theory is in no better position than HORT with respect to intimacy and misrepresentation.

2 The intimacy/misrepresentation objection

According to HORT, a distinct higher-order (HO) representation makes us conscious of our conscious states. But representational states by their nature can misrepresent their targets. This raises the following worry. Consider a case where the target of a HO representation is a red sensory state, but the HO representation represents to the effect that the subject is in a green sensory state. What will this experience be like, from the subject's point of view? According to HORT, the HO representation determines what it's like for the subject in conscious experience.[11] Therefore, a case of HO misrepresentation will be subjectively indistinguishable from a veridical case. In other words, in the misrepresentation case, it will seem to us that we are having a conscious green sensation, despite the fact that we are in a red sensory state.

But HORT allows a seemingly more radical kind of error. Alex Byrne writes,

> Suppose I have the higher-order [representation] that I am in a certain sensory state, and suppose that I'm not in that state ... there is only one answer to the question of what I will consciously experience: it will seem to me, phenomenologically, that I am in this sensory state (1997, p. 121).

Thus, it appears that the lower-order state isn't even needed for conscious experience. All that is required is the HO state.

This seems to open up HORT to a series of related objections. According to Byrne, the possibility of misrepresentation—particularly the more radical 'targetless' form—collapses the view's apparently relational explanatory structure. If there is something it is like for the subject when the HO state is present but the lower-order state is not, then the relational structure of the view is explanatorily idle. This, Byrne contends, leaves us wondering how a single representational state could explain consciousness, particularly conscious sensory experience.

Furthermore, the seeming irrelevancy of the first-order (FO) state, exposed by the possibility of misrepresentation, undermines the traditional philosophical notion of qualia. This aspect of the problem is stressed by Neander. She argues that any theory of consciousness ought to explain qualia, the "qualitative aspects of experience" (1998, p. 412). Qualia are the experienced properties of sensory states—for example, the conscious painfulness of pain or the conscious reddishness of a red sensation. They constitute conscious experience, according to Neander. But qualia, on HORT, seem to bifurcate—for example, pain and how pain consciously feels to us can apparently come apart. If we misrepresent a pain state as a tickle, it consciously feels to us that

[10] See Kriegel (2003a); Gennaro (2004, Sect. 2); Van Gulick (2004, pp. 71–74, 2006). See also Seager (1999).

[11] See, e.g., Rosenthal (1997, p. 744, 2004a, p. 40). Though Lycan is less happy to embrace this conclusion, it's clear it follows from his theory as well. See Lycan (1996, pp. 19–23, 1998, 2004). See also Armstrong (1963, 1981).

we're experiencing a tickle. Thus the pain state does not determine what it is like for the subject. But if qualia do not determine what it's like for the subject, then we are no longer talking about qualia "as traditionally understood" (1998, p. 420). Thus, the possibility of misrepresentation shows that HORT has changed the subject. It is no longer talking about qualia in the sense that we wanted explained at the beginning of our theorizing, and thus it isn't a theory of consciousness at all.

Joseph Levine expands on this point. Our reaction to the misrepresentation case, according to Levine, "seems to show . . . that one can't divorce the quality from awareness of the quality" (2001, p. 168). There is thus an 'intimate' connection between conscious qualities and our awareness of them which cannot be accounted for by HORT.[12] Our awareness of these qualities seems intrinsic to our conscious qualitative states, rather than a separate, dissociable thing. HORT must contend that this intrinsic connection is illusory. Further, as Neander stresses, the standard concept of qualia has it that qualia play a constitutive role in conscious experience, fixing what it is like for the subject. Levine writes that

> by forcing a division between the quale itself—the object of conscious awareness—and the awareness of it, the phenomenon we're interested in seems to disappear. Unfelt pains and unconscious reddish experiences may make some sort of theoretical sense if one already accepts an identification of qualia with neurophysiological states, but these states really do not seem to be the sorts of states we have in mind when thinking of qualia (2001, p. 168).

Qualia, as commonly understood, just *are* the 'feels' of conscious experience. And it certainly seems that they are central to the nature of consciousness. But on HORT qualia play "no genuine role in determining the qualitative character of experience" (2001, 108). Whereas qualia seem fundamentally intertwined with conscious experience, on HORT they need not even be present. Cases with and without the presence of qualitative states will be subjectively indistinguishable. We've clearly lost the intimate connection between qualitative states and our awareness of them in consciousness.

What's more, the possibility of misrepresentation seems to reveal a deeper incoherence in HORT. Kriegel (2003a) contends that the more radical form of misrepresentation forces the HORT theorist into an extremely counterintuitive position. According to HORT, conscious states are ones we are conscious of being in. But there is no first-order state present in this case, by hypothesis. Still, it will *seem* to the subject that she is in fact in a conscious state. There will be something it is like for her to have this experience—it will be subjectively indistinguishable from the veridical case. But what state is conscious in the targetless case? It can't be the HO state itself—we are in no way aware of that state. But the lower-order state isn't present at all. Surely, a state must be present for us to be in it. According to Kriegel, the HORT theorist is forced to conclude that "a person may be under the impression of being conscious when in fact she is unconscious" (2003b, p. 120). Thus, radical misrepresentation apparently exposes a theoretical incoherence in HORT—we can seem to be in conscious states when we are not. But if it seems any way to us at all, mustn't we be in a conscious state?

To summarize, the intimacy problem shows that, if HORT is correct, the qualitative aspects of conscious states—their sensory 'feels'—fail to play a 'genuine role' in conscious experience. At best, this exposes the explanatory shortcomings of the view; at

[12] The idea of 'intimacy' is stressed in Levine (2006).

worst, it shows that the view changes the subject or even falls into incoherence. We are left wondering how a single representational state could explain conscious experience. Further, we seem to have lost the phenomenon we wanted explained—the properties that constitute what it's like for a conscious subject. And HORT appears saddled with the claim that we can seem to be in conscious states when we are not. While this may not be an out-and-out contradiction in terms, it certainly does grave damage to our ordinary notion of consciousness. All this occurs because HORT allows the possibility of misrepresentation. SORT theorists craft their views with this in mind; indeed, the structure of SORT is in part determined by the need to account for intimacy.[13]

3 Same-order solutions, same-order problems

To avoid HORT's troubles with intimacy, SORT posits a single state to explain consciousness—a state that both represents the world and represents itself. The complex state is conscious *because* we are conscious of it, following the transitivity principle. And there is an intimate—indeed, a constitutive—relation between the qualitative features of consciousness and our awareness of these features. They are intertwined, intrinsic parts of one complex state. Furthermore, the complex state's identity conditions are defined by the proper binding of both elements—if one or the other is missing, then the conscious state does not exist. This apparently rules out cases of radical misrepresentation. And it apparently blocks Kriegel's worry that we might seem to be in a conscious state when we are not. For it to seem any way to us at all, a state that in part represents itself must be present; thus if we seem to be in a conscious state, the presence of a conscious state is ensured.

SORT theorists explain the nature of the complex conscious state in a variety of ways, and this provides somewhat different answers to the intimacy worry. Some hold that the complex conscious state is formed from two separate vehicles; one is about the world and the other is about the first state. Others hold that the complex conscious state is formed when the right sort of content emerges in a single vehicle; prior to that, no independent state possesses content that is about another state.[14] In what follows, I'll sketch the approaches of the three theorists mentioned above—Uriah Kriegel, Rocco Gennaro, and Robert Van Gulick—and I'll argue that each approach runs into trouble when confronting the problem. The particular problems together result in a trilemma for SORT: either the view is not significantly distinct from HORT, or it employs an unexplained (and arguably nonreductive) element, or it fails to conform

[13] It is important to note that, unsurprisingly, supporters of HORT do not accept that these dire consequences follow from the possibility of misrepresentation. Rosenthal, for example, has argued that the HO state fully accounts for what it is like for the subject in conscious experience, and that the worries presented all rely on an unsupported 'intrisincalist' intuition about conscious states. If we reject this intuition—that consciousness requires an intrinsic change in a conscious state, then the arguably worries have no bite. Rosenthal contends that a conscious state is one we are aware of ourselves as being in, and that this requires no intrinsic change in the state we are thus aware of. So long as we have an explanation of how we are aware, we will have an explanation of consciousness. See, in particular, Rosenthal (2004a). However, the SORT theorists at issue here all find Rosenthal's defense wanting, and turn to the SORT at least in part for that reason. Thanks to an anonymous referee for Synthese for pressing this point.

[14] See Levine (2006), for the 'one vehicle/two vehicle' way of typing same-order approaches. Kriegel and Gennaro defend 'two vehicle' views; Van Gulick defends a 'one vehicle' view. Thanks to an anonymous reviewer for Synthese for suggesting this clarification.

to the transitivity principle and collapses into a more standard functionalist theory. None of the options is appealing; so I conclude that SORT is in no better theoretical position than HORT.[15]

3.1 Kriegel's cross-order information integration model

Kriegel calls his version of SORT the 'cross-order information integration' (COII) model. The COII model explains subjective character—there being something it's like for the subject in conscious experience—in the following way:

> Consider a reddish experience of a circle ... The experience arises through the integration of information carried by two separate mental states, one representing the red circle, the other representing that very representation of the red circle. These two mental states ... have no subjective character, of course. But when their representational contents are integrated, a mental state arises which does have subjective character. The reason this third state has subjective character is that it has the right sort of representational content: it folds within it a representation of an external object and a representation of that representation (2005, p. 46).

The two integrated elements are both 'logical' parts of the complex conscious state; that is, the complex cannot exist without the parts, and furthermore, the proper 'binding' connection between the parts must occur. If any of these elements are lacking, the complex conscious state is not instantiated.[16]

The relationship between the parts, according to Kriegel, involves a 'psychologically real process' which integrates the elements; it's not enough that the two elements simply co-occur (2005, p. 32).[17] When the parts are properly integrated, we have a conscious experience with a particular character. Kriegel contends that there are plausible candidates for mechanisms of integration present in the brain. One possibility involves the kinds of neural processes posited in some theories to solve the 'binding problem,' the problem of explaining how information processed in separate regions of the brain is unified into a single experienced percept.[18] On this proposal, information is bound when the neural populations coding the information fire in synchrony. Kriegel writes,

> When two subpopulations of neurons in two disparate parts of the brain fire their electrical pulse in almost absolute synchrony (within the millisecond range), the

[15] See Levine (2006)(Sect. V), for a similar attack on SORT. Levine puts his objections in terms of intimacy and the explanatory gap and he focuses on Kriegel's version of SORT, but his critique generally makes the same point. Levine, of course, draws the stronger conclusion that no representational theory can close the explanatory gap. I hold the weaker claim that SORT is no better off than HORT. However, I am more optimistic that a representational approach can close the gap. See (reference suppressed).

[16] See Kriegel (2006) (Sect. 3). Kriegel uses the term 'logical part' to stress the fact that the identity conditions of the complex require the presence of the properly bound elements. I'm unsure about the role of 'logic' in the characterization, but I believe my gloss captures his intention—that a conscious state will not exist if all the parts are not present and properly bound.

[17] Kriegel employs 'psychologically real process' to distinguish his view from views like Gennaro's, which, according to Kriegel, only posit a notional connection between the higher-order and first-order elements of the complex state. According to Kriegel, the connection must be more than just *representation* of the world-directed element. See Kriegel (2005).

[18] See, for example, Crick and Koch (1990); von der Malsburg (1981).

> information they carry is bound into a single, cohesive content. That is ... the brain uses ... the *synchrony* of firing to represent the 'togetherness' (if you will) of these features as belonging to the same object (2005, fn. 69, emphasis in original).

It may be that the HO and FO components of a COII complex are bound by firing in synchrony. Kriegel further holds that the complex mental state "may misrepresent its own *properties*, but it is impossible that it should misrepresent its own *existence*" (2006, fn. 57, emphasis in original). This seemingly blocks radical misrepresentation, avoiding the case that causes the most serious problem for HORT. And, as noted above, the presence of a single complex state suggests a solution to the intimacy issue—conscious qualities and our awareness of them are intimately related because they are part of the same state.

However, the intimacy worry reemerges for COII.[19] The problem turns on what Neander calls 'the division of phenomenal labor' (1998). In Neander's terms, we need to know how the various elements present in COII do the phenomenal work—how they fix what it is like for the subject in conscious experience—if we are to judge whether intimacy is maintained. We can still ask what it will be like for a subject if the HO component misrepresents the subject as having a greenish experience, when the subject's FO component is a red sensory state. It seems that however the COII theorist responds there are problems.

Kriegel accepts the idea that a mental state is conscious when a subject is conscious of it—this is the transitivity principle. Further, Kriegel accepts that the HO component can misrepresent properties of the conscious complex, though he maintains that it cannot misrepresent the very existence of the complex it is a part of. As noted, this seems to give his version of SORT a distinct advantage over HORT: radical misrepresentation is ruled out. But the intimacy worry, as framed by Neander and Levine, turns on the role of the FO component in determining what it's like for the subject. And it's clear that in this respect, Kriegel's COII model has troubles with intimacy.

The HO component of a COII complex both determines that there is something it is like at all to be in a conscious state—it makes us aware of the state, in accordance with the transitivity principle—and what properties of that state the subject is aware of. Thus, if the HO component misrepresents a red FO component as a green FO component, it will seem to the subject that she is having a conscious experience of green. But this entails that what it's like for the subject varies independently of the FO component. The FO component in conscious perceptual experience contains a sensory element—this constitutes the difference between perceiving a red square and merely thinking of a red square. It follows that the qualitative properties of sensory states do *not* constitute what it is like for the subject in conscious experience. Thus, qualia, as defined by Neander and Levine, are not preserved on the COII. The view breaks the apparently intimate connection between qualia and our awareness of them, at least insofar as Neander and Levine are correct in how they characterize qualia. But leaving that issue to one side, it's clear that the COII is in no better shape than HORT regarding this issue.

[19] This point is stressed in Rosenthal (2004a, Sect. 5). In that section Rosenthal offers several additional arguments against SORT- most seriously, that it lacks a principled way of individuating mental states, undermining the claim that a single complex state accounts for consciousness. See also Kriegel (2003a).

Still, it does seem that Kriegel's COII avoids radical misrepresentation—misrepresentation by a HO state in the complete absence of a FO target. This alone appears to give his view a theoretical advantage over HORT. But this claim can be questioned. First, it's not clear that anything more than a linguistic advantage has been gained over HORT. If the HO component does all of the phenomenal work, what does it matter if a FO component is firing in synchrony with an erroneous HO representation? What it's like for the subject varies independently of qualia, showing that they have no 'genuine' role to play in conscious experience on the COII model.

Second, it's not clear that HO components can't 'misrepresent their own existence,' in the relevant sense. Consider the following example. On a sheet of paper, I compose the abstract to an article. I write, "In this very article, I show that such-and-such is the case, that God exists, etc." However, at publication, the abstract is attached to a completely different article. Question: Is the abstract about the new article it is erroneously attached to, or about the old article that it was written for? It seems clear that it's still about the old one. But what if I never wrote the old article at all, but merely wrote the abstract for an abstract-only conference submission? Then the abstract isn't about any article at all. Still, it is now attached to a new article, and has terms in it that seem to refer indexically—i.e. "this very article." It appears that even though it is 'integrated' with the new piece, the abstract doesn't refer to anything that exists—it refers only to the 'intentional inexistent' article I fully meant to write.

Why can't this sort of thing occur with Kriegel's COII? In sketching out the emergence of subjective character, Kriegel speaks of two separate states that then become components in an integrated conscious complex. With this in mind, imagine a HO state formed erroneously, in the complete absence of any target. Before being bound into a conscious complex, there is no reason it can't be about a merely intentional object. Now bind that state to a FO component. Regardless of the FO component's content, what it's like for the subject to be in this conscious complex will be fixed by the HO component, as argued above. We will be conscious of whatever the HO component is about. But the target of the HO component does not exist, ex hypothesi. Thus, we will seem to be in a state that does not exist—in Kriegel's terms, we will seem to be in a conscious state when we are not. This appears to be a case of radical misrepresentation, despite Kriegel's attempts to rule out such cases.

It follows that Kriegel needs the integration link to do substantial work in his theory to avoid the intimacy problem. Somehow, integration must allow the FO component to do real phenomenal labor—it must (at least in part) constitute conscious experience. Otherwise, the FO components—particularly the qualitative properties of FO components—will lack a genuine role in consciousness. But note that however integration works, it can't be cashed out in terms of representational content. Consider a subject who is in *unintegrated* states representing the world and representing the subject as being in those very states. That subject will not be in a conscious state, on Kriegel's view. But these separate states have the same *representational content* as a bound conscious complex—the only difference is integration. Thus, whatever difference integration makes, it won't be a representational difference. Integration plays some other sort of role.[20]

[20] There is an additional possibility that may seem to aid Kriegel here. Perhaps the HO component somehow alters the FO component when they are bound together, thereby creating a better match between the two. Then the FO component would have content intuitively better suited to play a role in phenomenal experience—we really would be in the state we represent ourselves as being in, easing the worry that the FO component is phenomenologically idle in consciousness. But this fails to

But it's not clear how to characterize this role in an informative way. If it's claimed that the complex can perform functions that the separate state cannot, we want a cataloging of those functions, and an explanation of why the separate states can't do the job. To simply say that the bound state plays the roles indicative of conscious states is question begging in this context. That is just to say that integration plays the consciousness role, not to provide an explanation of how integration achieves this feat in way that unintegrated states could not. Further, an explanation of integration must explain how the properties of the FO component become conscious even though we do not represent ourselves as being in states with those properties—this explanation is required if the FO component is to have a genuine role in the phenomenal labor. But that seems to require a mechanism independent of HO-representation, and thus independent of the transitivity principle. Such an approach is clearly out of step with the spirit of Kriegel's view, and raises the worry that the HO component will be rendered superfluous in an explanation of consciousness. If FO components are conscious already, why bring in the machinery of COII?[21]

What's more, the appeal to neural synchrony seems of little assistance here. In what way does knowledge that the integration is achieved by neural synchrony shed light on how the FO component achieves a genuine role in consciousness? All we thereby learn is that the states are now temporally coordinated—they are firing together. But knowledge of this neural fact fails to explain their *phenomenal integration*.[22] Further, as Kriegel acknowledges (2003a, pp. 491–492), binding is not sufficient for consciousness. There is good reason to hold that some nonconscious stimuli are bound; therefore, binding seems ill-equipped to explain how the FO component plays a genuine role in determining what it's like for a subject. For all we know, neural synchrony has nothing to do with consciousness per se.

And there is an additional worry. The type of binding Kriegel cites is involved in binding elements of a single percept—its color and shape, for example—not the binding of independent states. Synchronous firing seems to operate on a lower level, that of creating complex object representations, rather than at the level of mental states with complex 'sentence-sized' contents and mental attitudes. Thus, the mechanism for binding cited by Kriegel has nothing to say about the intimacy issue, which occurs at the level of states. In fairness, Kriegel introduces the neural binding suggestion as a mere possibility proof—a proof that it's at least possible that a neural mechanism might integrate conscious states. But be that as it may, we still want to know how binding provides a genuine role for the FO component in consciousness. Kriegel's mention of neural synchrony fails to explain this.

And there is a final danger lurking. As Kriegel also notes (2003a, p. 490), HORT offers only naturalistically 'kosher' ingredients to explain consciousness—both states at issue can be unpacked in representational or functional terms. But the binding mechanism needed by SORT looks to be something else entirely. Representational content won't account for the difference, and there does not seem to be a plausible

Footnote 20 continued
avoid the intimacy worry. First, we are still offered no explanation of how the FO component makes its phenomenal contribution. Second, we have no explanation of the mechanism creating the match in content. Finally, as long as the mere possibility of misrepresentation occurs—i.e., as long as the matching mechanism *may* go astray—the worry will emerge. Thus, the move only temporarily holds off the problem. Thanks to an anonymous referee for Synthese for pressing this point.

21 See Dretske (1995); Tye (1995, 2000). Cf. Neander (1998, pp. 430–434).

22 Cf. Levine (2006, Sect. V, pp. 192–193).

functional-role story to tell. But it's unclear what other naturalistically kosher ingredients are available. If it's not representation and it's not functional role, then it may just be the 'magic glue' of consciousness. But this, of course, gives up the central goal of a 'demystifying' explanation of consciousness.[23] I conclude that Kriegel's COII model still has troubles with intimacy, and therefore is no better off than HORT on this issue.

3.2 Gennaro's wide intrinsicality view

Gennaro calls his version of SORT the 'wide-intrinsicality view' (WIV) (1996, p. 23).[24] According to Gennaro,

> On the WIV, we have two parts of a single conscious state with one part directed at ('aware of') the other. In short, there is a complex conscious mental state with an inner, intrinsic relation between parts (2004, pp. 60–61).

In spelling out this 'intrinsic' relation, Gennaro cites the Kantian idea that concepts are presupposed in all experience.[25] He writes that "the understanding unconsciously 'synthesizes' the raw data of experience in order to produce the resulting conscious state" (2006, p. 237). The act of synthesis binds the HO conceptual element and the FO sensory element into one complex conscious mental state. Gennaro mentions that this process could be instantiated by neural 'feedback loops' of the sort described in Edelman and Tononi (2000), and that such loops are a ubiquitous feature of neural architecture (Gennaro, 2004, pp. 62–63).

In addition, Gennaro holds that the WIV is committed to a strong epistemic link between the FO and HO elements. He contends that

> there is ... a kind of infallibility between [the HO and FO parts] on the WIV, but this is not a problem. The impossibility of error in this case is merely within the complex CMS [conscious mental state], and not some kind of certainty that holds between one's CMS and the outer object (2006, p. 242–243).

Thus, whatever the synthesizing process amounts to, it precludes the possibility of representational error. Within the conscious complex itself, a mismatch between the HO and FO components cannot occur. If this proposal can be cashed out in acceptable terms, it will avoid the problem of misrepresentation altogether. And since the HO and FO elements of the complex state are both parts of the same state, there appears to be a good explanation of the intimate relation that holds between conscious qualities and our consciousness of them.

But if all we have on offer is that feedback loops in the brain are instantiated when the complex is instantiated, it is doubtful that Gennaro can make good on his infallibility claim. As Gennaro recognizes, feedback loops are an omnipresent feature of neural architecture, participating in both conscious and nonconscious processes. It's therefore unclear why such loops would have any particular affect on phenomenal

[23] This is the conclusion of Levine (2006). See note 15.

[24] Gennaro's version of SORT is in certain respects closest to HORT. In fact, he initially introduces his theory as a version of the higher-order-thought hypothesis, though it's clear that his view corresponds to the definition of SOMT given in Kriegel (2006) and to my characterization of SORT. Thanks to Rocco Gennaro for stressing this point and others in this section.

[25] See, for example, Gennaro (1996, Chap. 3, pp. 44–57, 2004, p. 62, 2006, Sect. 4.1).

experience. Further, we are given no inkling at all about how such loops are supposed to ensure a perfect match between the FO and HO components of the complex state. What natural process could provide such a guarantee? Armstrong's famous argument against the infallibility of introspection turns simply on the idea that when we are dealing with causal, mechanistic processes, it's always possible that something can go wrong.[26] This applies equally well to whatever mechanisms might be involved in Gennaro's WIV. Given the ubiquity of feedback loops, why couldn't there be a mismatch between the content of the HO and FO components? Why, for example, couldn't a HO representation of a red sensation engage in a feedback loop with a green sensory state? That the system might correct itself quickly is beside the point. The mismatch could occur, undermining the infallibility claim. So long as ordinary neural mechanisms are involved, error can't be ruled out. So a guarantee of infallibility arguably blocks the WIV from a demystifying explanation of consciousness.

But we can step back from Gennaro's strong infallibility claim—this idea alone seems to threaten the naturalistic credentials of the view.[27] However, then we are back to the initial worry brought on by misrepresentation. The division of phenomenal labor is again the central issue. If what it's like for the subject is dictated by the HO component alone, then the troubles with intimacy reemerge—what it's like for the subject varies independently of the qualitative properties of our sensory states, blocking qualia from a genuine role in consciousness. Thus, Gennaro, like Kriegel, needs to explain how the FO component does phenomenal work. In this context Gennaro cites Kant as a source of possible enlightenment. Kant argued that experience arises out of the synthesizing of concepts with sensations—concepts 'operate on' sensations to make experience possible. Gennaro argues that the complex conscious state can be seen in this light. The HO component is conceptual in nature—it involves the application of concepts to one's own sensory experience. And the FO component includes sensory qualities, at least in perceptual experience. Thus, we can view the complex conscious state as the product of a synthesis between FO sensations and HO concepts. The two combine to create the conscious state.

But it's not fully clear what 'synthesis' amounts to here. One concern is that Kant's arguments establish something about the necessary conditions for conscious experience to be *possible*, rather than telling us about how such a mechanism works in the actual world.[28] Another is that the most plausible way to cash out synthesis in a naturalistic theory of mind is in terms of representation. We synthesize our concepts and sensations by representing ourselves, in HO representation, as being in sensory states. I conceptualize myself as being in various sensory states, and thus synthesize my concepts and sensations. However, as noted in connection with Kriegel's view, this fails to provide the intimate connection needed to block the worry. Merely representing our sensory states fails to provide the needed intimacy.

26 Cf. Ayer (1956, Chap. 1, p. 29).

27 And indeed, this may well be stronger than Gennaro himself intends (see Gennaro (2006), Sect. 4.6). Perhaps all he intends is that we cannot be in error that we are in a conscious state—a claim parallel to Kriegel's assertion that we can misrepresent the properties of our conscious state, but not its existence. If this is the proper reading, then the foregoing criticisms are beside the point. But if the stronger reading is intended, then it is reasonable to ask how any *natural* process can guarantee the match. Given Armstrong's and Ayer's point, an answer to this question seems unlikely, rendering the connection mysterious, from the point of view of naturalism.

28 Thanks to David Rosenthal for suggesting this point, and for helpful comments throughout.

And again, the invocation of feedback loops fails to offer any explanatory help. Feedback loops only provide a causal connection between the HO and FO components; they tell us nothing about which elements do phenomenal work, or what happens in cases of mismatch. If we do not have a clear, prior idea of how synthesis works to ensure a match between the components, learning that the process is instantiated by feedback loops tells us little.

Furthermore, there appear to be nonconscious cases of synthesis—in subliminal perception and masked priming, for example.[29] Cases of masked priming and subliminal perception both demonstrably involve concepts. That we subliminally perceive a stimulus *as* money explains how we disambiguate the word "bank" in masked priming experiments. Subjects report no awareness of the priming image, yet behave in ways that clearly indicate that the image was processed and categorized. This indicates the synthesis of concepts and sensations in the absence of conscious experience. So synthesizing may have nothing to do with consciousness per se; rather, it may be a condition on perception itself, rather than consciousness.

This raises the question of how synthesis could enable a genuine role for the FO component in conscious experience. If synthesis alone isn't sufficient for consciousness, it isn't clear how it can do the job in conjunction with HO representation. Consider once more a case where the HO component misrepresents a FO component. For the FO component to do real phenomenal work, some of its properties must be conscious independently of how they are represented by the HO component—otherwise, the HO component does all the phenomenal work. But if synthesis doesn't make states that we're unaware of being in conscious—as in cases of synthesized subliminal perception—why would it make the unrepresented properties of a FO component conscious? We're not aware of those properties by HO representation, ex hypothesi, so synthesis alone must make the properties of the FO component conscious. We are left wondering what the process of synthesis amounts to. It is not enough to say that synthesis is a 'combining' or that in synthesis 'concepts operate on sensations'—that fact is agreed on. And it may be that Kant is correct in his arguments that experience requires the synthesis of concepts and sensations. But we want to know, in nonmysterious terms, how this process works, in a way that ensures intimacy. In the absence of such an explanation, we are left again, so it seems, with 'magic glue.' Either the connection is representational, in which case intimacy is still a problem; or it's something else, in which case we are in danger of losing our nonmysterious explanation.

A final point concerning my explanatory demands on WIV and SORT in general. I am not demanding a solution to the so-called 'hard problem' of consciousness—the problem of explaining why any mental state is conscious at all. Rather, I am asking for an explanation of the mechanism that accounts for our awareness of our conscious states—the phenomenon fixed by the transitivity principle. Given this explanandum, explicitly endorsed by Gennaro and the other SORT theorists considered here, we can ask how that awareness is achieved. Gennaro refers to synthesis; my worry is that either synthesis fails to distinguish the WIV from HORT, or synthesis can't be cashed out nonmysteriously. It is an independent issue if solving these problems, and thus explaining the transitivity principle, would thereby solve the hard problem. I believe it would, but that is not argued for here.

[29] On subliminal perception, see Merikle, Smilek, and Eastwood (2001). On masked priming, see Marcel (1983a, b).

3.3 Van Gulick's HOGS model

Van Gulick calls his approach the 'higher-order global states' model (HOGS). The model differs from the other versions of SORT discussed so far in endorsing the main tenet of 'global workspace' theories of consciousness—the idea that conscious states are 'globally available,' accessible to a wide range of mental systems and processes.[30] However, Van Gulick distinguishes his view from more standard workspace theories by arguing that there is a crucial element of self-awareness in consciousness—that we are always to some extent aware of our conscious states. This brings his view in line with the other SORT views detailed here.

As noted in the introduction, Van Gulick registers his approval of the transitivity principle. In addition, he also cites a number of phenomenological considerations suggesting a close connection between self-awareness and consciousness. For example, he contends that conscious experience involves the integration of the self into a unified phenomenal world—to experience a world is to have a perspective, and a perspective seemingly requires a self. Furthermore, our conscious experience flows in ways that we can control and alter—we mold experience by our deliberate interactions with the world. But at the same time, we are altered and affected by our perceptual environment. This suggests that the world- and self-awareness are of a piece—the presence of one shapes the presence of the other. Finally, the objects we phenomenally experience are imbued with qualities that reflect our understanding and know-how concerning those objects. A paperweight looks to have a substantial heft and presence because of my knowledge of how to interact with it. The appearance of our phenomenal world thus seems shot-through with self-awareness, according to Van Gulick. This suggests that conscious experience fundamentally involves a background of self-awareness.

Bringing these ideas together gives us the HOGS model. Consciousness, according to the model, is marked by globally accessible states which contain an element of reflexive self-awareness. But Van Gulick contends that self-awareness does not require explicit representation. We can be aware of our globally accessible states without forming distinct representations of them. Van Gulick thus writes,

> The basic idea of . . . the HOGS . . . model of consciousness is that transforming a nonconscious state into a conscious one is a process of *recruiting* it into a *globally integrated complex* whose organization and intentional content embodies a heightened degree of *reflexive self-awareness*. The meta-intentional content on the HOGS model is realized not by a distinct and separate external meta-state but rather by the organization of the complex global state itself, and the object state is a component of the global state into which it has been recruited (2006, p.24, emphasis in original).

Van Gulick holds that a wide range of behavior entails 'implicit' self-awareness. Implicit self-awareness does not require a representational state that is about the organism or its mental states as such. Rather, the awareness is contained or embodied in the structures that underwrite certain types of complex behavior—behavior involving interactive feedback between organism and environment or between

[30] Van Gulick holds that "conscious mental states . . . are more widely and powerfully influential within the minds in which they occur. They are more broadly and richly connected, and their contents are more widely accessible by other subsystems throughout their containing mind-brain" (2006, p.24). See also Van Gulick (2004, p. 75ff). For the global workspace theory, see Dennett (1991, 2001); Baars (1998, 1997), Baars, Ramsøy and Laureys (2003).

systems within an organism. Implicit self-awareness is present in many basic learning and adaptive motor processes, according to Van Gulick. And, to the point, it is present to a high degree in the globally accessible states instantiating consciousness on the HOGS model, suggesting an explanation of our awareness of our conscious states–we are implicitly aware of being in those states.[31]

We are left with a single, complex state that is both about the world and (implicitly, at least) about itself. Thus, Van Gulick's HOGS model is a version of SORT. How, then, does the HOGS model deal with intimacy and misrepresentation? At first glance, it appears to handle such cases effectively. The FO component itself becomes conscious not by being explicitly represented, but by being integrated into a global state. Thus, the FO state itself is conscious, in virtue of the new connections it has to other states and processes. And since the HO component is implicit, it's not clear how it could come apart from the FO component in a way that allows a mismatch. Our awareness of our conscious states is embodied by the connections those states have to other states and processes—there is no distinct HO state to get things wrong. The worry is thus avoided. Further, the connection between conscious states and our awareness of them appears intimate—the awareness is seemingly constituted by the global processes that instantiates consciousness. Thus, qualitative states play a genuine role in consciousness.

But despite initial appearances, the HOGS model does not save SORT. The HOGS model may avoid the problems of intimacy and misrepresentation, but at the cost of failing to explain the transitivity principle. Representational theories, as I've defined them here, use representation to explain how we are conscious of our conscious states. This, as Van Gulick notes, offers a route to a demystifying explanation of consciousness. Further, it fits well with common sense—our ordinary notion of awareness arguably involves representation. But the commonsense notion of awareness involves *explicit* representation. If I perceive or think about something—our ordinary ways of being aware—I have a perception or a thought explicitly informing me that the thing is present. If the presence of that thing is merely implied by my behavior, we wouldn't intuitively conclude that I was aware of it. Unless there is an occurent thought or perception guiding my behavior, awareness is intuitively lacking. Even in cases of so-called unconscious perception—like that of Armstrong's long-distance truck driver—plausibly involve explicit representation by the subject. If there is no active representational state controlling her actions, how does the truck driver avoid running off the road? The best explanation, from our everyday perspective, is that a representation that explicitly represents the road is present and occurently active in the truck driver. One might argue that the representation must be conscious; however, it seems clear either way that the representation must explicitly represent the road if it is to do its job.[32] Our commonsense notion of awareness, therefore, even in the case of nonconscious perception, involves explicit representation. By rejecting explicit representation, the HOGS model loses the most straightforward means of explaining the transitivity principle. Thus, something else must explain how we are conscious of our conscious states.[33]

[31] Van Gulick often uses the terms 'implicit self-understanding' for the self-reflexive element in HOGS. In keeping with the focus of this paper, I use the term 'implicit self-awareness' to highlight the role this element must play if it is to explain the transitivity principle. See Van Gulick (2004, Sect. 3, 2006, Sect. II).

[32] Cf. Carruthers (2000, Sect. 6.1 and 6.2).

[33] It should be noted that a number of approaches reject the idea that *explicit* representation is required for awareness of our conscious states. The phenomenological approach, for example,

But global accessibility is ill-suited to fill that role. Global accessibility is characterized in causal-informational terms distant from our ordinary notions of consciousness and awareness. A globally accessible state is one whose informational content can be accessed by a variety of systems and processes. But this tells us nothing about a subject's awareness of the state. Further, the claim that global accessibility and awareness are correlated tells us nothing about *why* the awareness occurs. In addition, global accessibility is a dispositional notion. States are globally accessible when they are available to systems, or 'poised' for use. But the awareness we have of our conscious states seems occurent and categorical—it's not clear how a disposition to be accessed could account for this.[34] The virtue of the representational approach is that it provides a clear connection to our ordinary notions of awareness and consciousness. In addition, it employs an occurent, categorical process rather than a dispositional one. Global accessibility, on the other hand, falls prey to the usual worries that plague most functionalist accounts of consciousness.[35] Simply drawing extra lines connecting one of the boxes in a flow chart can't tell us why we're therefore aware of the contents of that box. Something more is clearly needed.

And, of course, Van Gulick attempts to provide the added ingredient. Implicit self-awareness (ISA) is allegedly present to a high degree in HOGS, and this, it may seem, makes us conscious of our conscious states. However, the notion of ISA, as presented by Van Gulick, falls short as an explanation of the transitivity principle. As argued above, it is disconnected from our ordinary notions of awareness and consciousness—those notions involve explicit awareness. But this may seem indecisive; after all, how much theoretical work can our commonsense notions do here? However, there are plausible cases of nonconscious states that possess a high degree of ISA. This entails that ISA is not sufficient to explain consciousness as characterized by the transitivity principle, even if we waive the worry about our ordinary notion of awareness.

Van Gulick contends that processes involving ISA go on all over the phylogenetic scale, and at a wide variety of levels of organization within a single creature. For example, he holds that ISA is present in the mechanisms in rats that underwrite a learned aversion to noxious foods. Van Gulick explains that the learning mechanisms involved in this process

> are 'tuned' in at least two respects: to the causal structure of the external world and to the mental or intentional structure of the organism's own internal

Footnote 33 continued
generally holds that our awareness of our conscious states is not by way of ordinary representation, and instead employs a nonexplicit sort of awareness. But this arguably is different from our everyday notion, and phenomenological theorists concede as much. Phenomenology demands an expert evaluation of conscious experience, in order to bracket the irrelevant elements. Since only a trained phenomenologist can perform such an investigation, the results plausibly go beyond our everyday conception. Levine (2006) also holds that our awareness of our conscious states is not by way of ordinary representation. Instead, it is a sui generis sort of awareness. However, Levine holds that the challenge of explaining this awareness is part of what leads to the explanatory gap. Thus, according to Levine, rejecting explicit representation leaves one without an explanation. My point here is only that it requires more than simply stating that implicit representation occurs in consciousness. We need an explanation of why this makes us aware of our conscious states, and how that process works, in naturalistic terms. Thanks to an anonymous referee for pressing this issue.

34 Rosenthal (2002b), Seager (2001), and Kriegel (2005), among others, stress this point.

35 Namely, that functional processes are ubiquitous in nature, making it unclear how function alone could account for the distinction between conscious and nonconscious states; and that functional notions are dispositional, rendering them ill-equipped to explain the apparently occurent nature of conscious experience. See, e.g., Block (1980, 1995), Rosenthal (2002b).

> organization. Its job indeed is to better harmonize the two in the context of the individual organism's local situation (2006, p.22).

This sort of integrated feedback-driven process is constitutive of ISA. Van Gulick notes that 'self-reflexive loops' of this nature are a central building block in the mind's evolution. The coordination and 'harmonization' that this sort of organizational structure engenders is a foundational feature of mentality.

But this suggests that many ISA processes probably occur independently of the organism's awareness. It is controversial at best to think that rats are in any intuitive sense aware of the processes that realize their aversion responses. Further, we are rarely conscious of the multitude of feedback-driven loops and connections present in our psychology. For example, we learn to walk, to play games, and to speak—all processes involving a high degree of ISA. But this does not necessarily make us conscious of the on-going states instantiating these achievements. We nonconsciously learn the complex series of motor controls needed for walking, the rich collection of social regulations implicit in children's games, and the recursive rules of natural language grammar, all without such states becoming conscious. We clearly must monitor, update, and 'harmonize' our mental states in these tasks in the ways indicative of ISA—we clearly have the requisite know-how or understanding. Thus, it appears that ISA is not sufficient to make us conscious of our states. We can be implicitly self-aware of a state, and not be conscious of the state in any intuitive sense.

ISA simply characterizes the organization needed in tasks involving feedback between the organism and the environment or between various subsystems of the organism. It is present in a wide-range of processes that occur independently of our awareness. But Van Gulick can respond that the required level of ISA—enough to make us conscious of our states—is only present in richly interconnected HOGS. HOGS are marked by their high degree of ISA, owing to their many connections to psychological systems and processes. The reason we are not conscious of the processes I mentioned above is because they just don't have the needed degree of ISA.

But there are two problems with this suggestion. One is that there are plausible cases of *globally accessible*, implicitly self-aware states that we are nonetheless not conscious of in any intuitive sense. The other is that even if there were no such cases, we still would lack an explanation of what the connection is between ISA and our awareness of conscious states. Regarding the first point, consider a situation where I am jealous, but unaware that I'm jealous. I may fancy myself above such petty emotions or too enlightened to succumb to base instincts, so I suppress my feelings, and even sincerely deny that I'm jealous when asked. But it will be apparent—especially to those close to me—that I am jealous by the way I act. I'll act rudely towards the person I'm jealous of, I'll misinterpret his words and actions, I'll behave aggressively towards him, and so on. Later, I might become conscious of my jealousy, but at the time of the confrontation, I will not be conscious of it.

My jealousy in this case certainly appears globally accessible. It controls my moods, my other emotions, my judgments and my perceptions involving the target of my jealousy. It even affects physiological reactions like my temperature and my rate of heartbeat and respiration. The state is not only available to a wide range of systems and processes; it is actively accessed by many of them. Furthermore, I am clearly implicitly self-aware of the state, in Van Gulick's sense. My jealousy shapes my interactions with my social environment. It determines my judgments, my perceptions,

and my behavioral reactions, even my speech. And this in turns feeds back onto my emotional state, affecting its evolution. This sort of interaction is the mark of ISA.

Further, the process of repressing my jealous state involves ISA as well. I must continually recognize and track the state, in order to keep it from becoming conscious. And I must construct an alternative interpretation of my feelings even as I continue to nonconsciously display my jealousy. Surely this involves a considerable degree of ISA. But the entire process can occur without my being aware of it. Therefore, it appears that being a globally accessible state with a high degree of ISA is insufficient to make us aware of that state.[36]

In addition, there are more intellectually oriented cases of globally available ISA states of which we are unaware. Consider an instance of nonconscious problem solving. A philosophical problem may occupy much of my mental energy and demand the implicit, harmonizing know-how of ISA for its solution. Still, I might come to the solution 'out of the blue,' indicating that the problem solving occurred without my being aware of it. And depending on the nature of the problem, a wide range of systems might be involved in its solution. Finally, there may even be candidates for globally accessible ISA sensory states that we aren't conscious of being in. Visual stimuli in masked priming and subliminal perception tasks can affect a wide variety of subsystems—they can influence patterns of thought and behavior, and they can influence emotional and value judgments. Subliminally perceived odors can affect how we judge a potential mate. Chronic pain can affect mood, focus, intellectual abilities, and so on, even though it may flit in and out of our consciousness. While such cases need not engage all of our mental systems at one time, they seem to be available to a wide range of processes.[37] And this entails that they possess a high degree of ISA—they can enter into a range of process involving interactive, harmonizing know-how. Again, it seems that HOGS can occur without our being aware of them.

And even if we could somehow rule out all these cases as nonconscious, we'd still be left wondering how an implicit structural property, one that is often present in uncontroversially nonconscious processes, makes us aware of anything. Van Gulick does, however, cite the rich intertwining of world and subject that marks our phenomenal experience. Consciousness itself seems shot-through with a background of self-awareness. Does this point to HOGS, allowing us to reason from phenomenal appearances back to underlying processes? For purposes of argument, we can grant

[36] One might contend that I've left out a vital component of the global processes required for consciousness on the HOGS model, namely the *reportability* of globally accessible conscious states. In processes where states are reportable, there do not seem to be cases of nonconscious globally accessible states, states that will inter alia contain a high degree of ISA. However, we are left wondering why accessibility to the reporting system could make this difference on the HOGS model. One possibility is that to report, we must explicitly represent the state we are in—this plausibly explains the content of the reporting speech act (see, e.g., Rosenthal, 2002a). But this move is not open to the HOGS theorist, who holds that implicit representation is crucial to consciousness. Further, reportability is a dispositional notion, ill-equipped to explain the occurent appearance of consciousness. Finally, it is unclear how reportability helps to explain the transitivity principle—why would being disposed to report make us aware of our conscious states, and why would its lack block such awareness? This highlights both the troubles with explaining consciousness in terms of reportability, and the impotence of ISA in explaining the transitivity principle. We are left wondering why the disposition to report makes our states conscious, and why we are *unaware* of the rich but unreportable states suggested by the jealousy example. Thanks to an anonymous reviewer for Synthese for pressing this point.

[37] See, e.g., Dell' Acqua and Grainger (1999), Winkielman, Berridge, and Whilbarger (2005), Mitchell, Kahn, and Knasko (1995), Richardson (1996), Hart, Wade, and Martelli (2003), Grisart and van Der Linden (2001).

that experience has all the robust features of self and world Van Gulick describes. But even so, it remains unclear that the HOGS model offers an explanation of the aspects of phenomenal experience that are at issue. If global accessibility and ISA are insufficient for our consciousness of our conscious states, then we still do not know why we are conscious of this rich phenomenal world, even if it's true that that world contains a necessary element of meta-intentionality. Global accessibility offers no explanation of this awareness, and ISA doesn't intuitively make us aware of anything—ISA is present even when such awareness is lacking. Global accessibility and ISA still fail to explain the transitivity principle.

One plausible means of explaining the insight is, of course, explicit representation. It explains why some globally available processes are conscious and others are not. Further, it accounts for the alteration in self-awareness that makes us conscious of our conscious states. But this is to walk back into the problems that the HOGS model seeks to avoid. If it's explicit self-representation, we can again ask about what occurs when there is a mismatch between the self-representation and its target. By making the self-awareness implicit, the HOGS model avoids this worry—if nothing is represented, we cannot misrepresent. But it does so at the cost of explaining how we are conscious of our conscious states. We can conclude that the HOGS model is not a representational theory. It thereby loses the allegedly demystifying benefits of such an approach; furthermore, it inherits the troubles plaguing more straightforward functionalist views, namely the ubiquity of functional processes in nature and their dispositional as opposed to occurent character.[38,39]

4 Conclusion

To wrap up, SORT fails to avoid the intimacy/misrepresentation worry. The view either employs representation to connect the HO and FO elements of the complex conscious state, in which case it simply moves the worry to a new location: inside the state itself. This still fails to deliver a genuine role for the FO component. Or the view employs a new relation to connect the two elements; however, there seems to be no good naturalistic candidates at hand. Or the view slides back into the functionalist camp, and therefore fails to explain our awareness of our conscious states. In the end, and despite the theoretical richness of these three models, SORT is not in a better explanatory position than HORT. Both views, it seems, have trouble with intimacy.

[38] Indeed, Van Gulick does not use the term 'representation' in describing his view. But he does cite the higher-order approach as a motivation for the HOGS model, and he contends that the model explains the sense of 'consciousness' captured by the transitivity principle (see Van Gulick, 2006, Sect. V). Further, his view fits well with the characterization of SOMT given in Kriegel (2006), and SORT as I've defined it here.

[39] Again, note that I am not demanding that the HOGS model solve the 'hard problem' of consciousness, per se (see Chalmers, 1996). That is, I am not asking for an explanation of why conscious qualitative states possess the particular conscious quality they do, nor how it is that physical states can be conscious at all. I am instead demanding an explanation of the transitivity principle—how we are conscious of our conscious states. This is certainly one of Van Gulick's explanatory goals, even though he's doubtful that's all there is to explaining consciousness (see Van Gulick, 2006, Sect. V). I am more optimistic that an explanation of the transitivity principle does answer the hard problem, but in this context, I am only arguing that the HOGS model fails to explain the transitivity principle. It a fortoriori fails to explain the hard problem, if, pace Van Gulick, an explanation of the transitivity principle solves that problem.

Acknowledgements A version of this paper was presented at the 'Self-Representational Approaches to Consciousness' conference, Center for Consciousness Studies, University of Arizona, March 18th–20th, 2005. I am grateful to the conference participants for helpful feedback. Thanks to Rocco Gennaro, Uriah Kriegel, David Rosenthal, and an anonymous reviewer for Synthese for helpful comments on an earlier draft.

References

Armstrong, D. M. (1963). Is introspective knowledge incorrigible? *Philosophical Review, 62*, 417–432.
Armstrong, D. M. (1968). *A materialist theory of the mind*. New York: Humanities Press.
Armstrong, D. M. (1981). *The nature of mind and other essays*. Ithaca Press: Cornell University.
Ayer, A. J. (1956). *The problem of knowledge*. London: Macmillan.
Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. Oxford: Oxford University Press.
Baars, B. J., Ramsøy, T. Z., & Laureys, S. (2003). Brain, conscious experience, and the observing self. *Trends in Neuroscience, 26*(12), 671–675.
Block, N. J. (1980). Troubles with functionalism. In N. Block (Ed.), *Readings in the philosophy of psychology* (Vol. 1). Cambridge, MA: Harvard University Press.
Block, N. J. (1995). On a confusion about a function of consciousness, *Behavioral and Brain Sciences, 18*(2), reprinted in N. J. Block, O. Flanagan, & G. Güzeldere (Eds.). (1997), *The nature of consciousness: Philosophical debates* (pp. 375–416). Cambridge, MA: The MIT Press.
Byrne, A. (1997). Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies, 86*(2), 103–129.
Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge University: Cambridge Press.
Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford and New York: Oxford University Press.
Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences, 2*, 263–275.
Dell' Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition, 73*, B1–B15.
Dennett, D. C. (1991). *Consciousness explained*. Boston: Little Brown.
Dennett, D. C. (2001) Are we explaining consciousness yet? *Cognition, 79*, 221–237.
Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: The MIT Press.
Edelman, G., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.
Gennaro, R. J. (1996). *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness*. Amsterdam and Philadelphia: John Benjamins Publishers.
Gennaro, R. J. (2004) Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and levine. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 45–66). Amsterdam and Philadelphia: John Benjamins Publishers.
Gennaro, R. J. (2005), The HOT theory of consciousness: Between a rock and a hard place? *Journal of Consciousness Studies, 12*(2), 3–21.
Gennaro, R. J. (2006). Between pure self-referentialism and the (extrinsic) HOT theory of consciousness. In U. Kriegel, & K. Williford (Eds.), *Consciousness and self-reference*. (pp. 221–248). Cambridge, MA: MIT Press/Bradford Books.
Grisart, J. M., & Van Der Linden, M. (2001). Conscious and automatic uses of memory in chronic pain patients. *Pain, 94*, 305–313.
Hart, R. P., Wade, J. B., & Martelli, M. F. (2003). Cognitive impairment in patients with chronic pain: The significance of stress. *Current Pain and Headache Reports, 7*, 116–126.
Kriegel, U. (2003a). Consciousness, higher-order content, and the individuation of vehicles. *Synthese, 134*, 477–504.
Kriegel, U. (2003b). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy, 33*, 103–132.
Kriegel, U. (2006). The same-order monitoring theory of consciousness. In U. Kriegel, & K. Williford (Eds.), *Consciousness and self-reference*. (pp. 143–170). Cambridge, MA: MIT Press/Bradford Books.

Kriegel, U. (2005). Naturalizing subjective character. *Philosophy and Phenomenological Research*, *71*(1), 23–57.
Levine, J. (2001). *Purple haze: The puzzle of consciousness*. New York: Oxford University Press.
Levine, J. (2006). Conscious awareness and (self-)representation. In U. Kriegel, & K. Williford (Eds.), *Consciousness and self-reference*. (pp. 173–197). Cambridge, MA: MIT Press/Bradford Books.
Lycan, W. G. (1987). *Consciousness*. Cambridge, MA: The MIT Press.
Lycan, W. G. (1996). *Consciousness and experience*. Cambridge, MA: The MIT Press.
Lycan, W. G. (1998). In defense of the representational theory of qualia (replies to Neander, Rey and Tye). In J. E. Tomberlin (Ed.), *Language, mind, and ontology: Philosophical perspectives (Vol. 12)*. Atascadero, CA: Ridgeview Publishing Company.
Lycan, W. G. (2001). A simple argument for the higher-order representation theory of consciousness. *Analysis, 61*(1), 3–4.
Lycan, W. G. (2004). The superiority of HOP to HOT. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness*. Amsterdam and Philadelphia: John Benjamins Publishers.
Marcel, A. J. (1983a). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15*, 197–237.
Marcel, A. J. (1983b) Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15*, 238–300.
Merikle, P. M., Smilek, D., & Eastwood, J. D. (2001). Perception without awareness: Perspectives from cognitive psychology. *Cognition, 79*, 115–134.
Mitchell, D. J., Kahn, B. E., & Knasko, S. C. (1995). There's something in the air: Effects of congruent or incongruent ambient odor on consumer decision making. *Journal of Consumer Research, 22*(2), 229–238.
Neander, K. (1998). The division of phenomenal labor: A problem for representational theories of consciousness. In J. E. Tomberlin (Ed.), *Philosophical perspectives 12: Language mind and ontology* (pp. 411–434). Boston: Blackwell Publishers.
Richardson, S. (1996). Scent of a man. *Discover, 17*(2), 26.
Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49*(3), 329–359.
Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 729–753). Cambridge, MA: The MIT Press.
Rosenthal, D. M. (2000). Consciousness and metacognition. In D. Sperber (Ed.), *Metarepresentation: Proceedings of the tenth vancouver cognitive science conference*. New York: Oxford University Press.
Rosenthal, D. M. (2002a). Explaining consciousness. In D. Chalmers (Ed.), *Philosophy of mind: contemporary readings*. Oxford: Oxford University Press.
Rosenthal, D. M. (2002b). How many kinds of consciousness. *Consciousness and cognition, 11*(4), 653–665.
Rosenthal, D. M. (2004a). Varieties of higher-order theory. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 19–44). Amsterdam and Philadelphia: John Benjamins Publishers.
Rosenthal, D. M. (2004b). Being conscious of ourselves. *Monist, 87*(2), 161–184.
Seager, W. (1999). *Theories of consciousness: An introduction and assessment*. London and New York: Routledge.
Seager, W. (2001). On dispositional HOT theories of consciousness. *SWIF philosophy of mind*, http://www.swif.uniba.it/lei/mind/forums/seager.htm.
Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: The MIT Press.
Tye, M. (2000). *Color, content, and consciousness*, Cambridge, MA: The MIT Press.
Van Gulick, R. (2001). Inward and upward: Reflection, introspection, and self-awareness. *Philosophical Topics, 28*, 275–305.
Van Gulick, R. (2004). Higher-order global states (HOGS): An alternative higher-order model of consciousness. In R. J. Gennaro (Ed.), *Higher-order theories of consciousness* (pp. 67–92). Amsterdam and Philadelphia: John Benjamins Publishers.
Van Gulick, R. (2006). Mirror, mirror—Is that all? In U. Kriegel, & K. Williford (Eds.), *Consciousness and self-reference*. (pp. 11–39). Cambridge MA: MIT Press/Bradford Books.
von der Malsburg, C. (1981). The correlation theory of brain function. Technical Report 81-2, Max-Planck-Institute for Biophysical Chemistry, Göttingen.
Winkielman, P., Berridge, K. C., & Wilbarger, J. L. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality and Social Psychology Bulletin, 31*(1), 121–135.