Between Pure Self-Referentialism and the (Extrinsic) HOT Theory of Consciousness

Rocco J. Gennaro

[Forthcoming in Kriegel/Williford (eds.), Consciousness and Self-Reference, MIT Press]

1. Introduction: Three Views of State Consciousness

The notion that there is a self-referential aspect to conscious mental states has a long tradition, going back as far as Aristotle (Caston 2002) and, more recently, Franz Brentano (1874/1973: 153) who famously held that “...every mental act...includes within it a consciousness of itself. Therefore, every mental act, no matter how simple, has a double object, a primary and secondary object.”¹ For those of us who wish to offer a reductive representational account of consciousness in mentalistic terms, this idea may hold the key to a successful explanation of state consciousness.² In the context of arguing for the thesis that consciousness entails self-consciousness (Gennaro 1996), I introduced the “wide intrinsicality view” (WIV) whereby conscious mental states do indeed contain a crucial self-referential element. On the WIV, first-order or world-directed conscious mental states are complex (or global) states such that one part of the state is directed at another part. Thus, conscious states are to be individuated widely and consciousness is intrinsic to them. Although similar in many ways to David Rosenthal’s (extrinsic) higher-order thought (EHOT) theory, there are some key differences, as will become clear in the next section.³ In section 3, I criticize what we might call Brentano’s “pure self-referentialism” (PSR); namely, that a conscious mental state is literally directed back at itself. I argue against PSR and show that the WIV is indeed a more plausible theory of state consciousness. As we shall see, the WIV is located importantly between PSR and EHOT theory, and something at least similar to it can be found in the more recent literature.⁴ Finally, in section 4, I further clarify the WIV by responding to a number of important objections.⁵

We thus have three positions with respect to first-order world-directed conscious states:

**EHOT**: A mental state M of a subject S is conscious if and only if S has a distinct
(unconscious) mental state $M^*$ (= a HOT) which is an appropriate representation of $M$.

**WIV:** A mental state $M$ of a subject $S$ is conscious if and only if $S$ has a suitable (unconscious) meta-psychological thought, $M^*$ (= MET), directed at $M$, such that both $M$ and $M^*$ (= MET) are *proper parts of* a complex conscious mental state, CMS.$^6$

**PSR:** A mental state $M$ of a subject $S$ is conscious if and only if $S$ has a mental state $M^*$ which is an appropriate representation of $M$, and $M = M^*$. $^7$

All three views take very seriously the intuitive notion that a conscious mental state $M$ is a state that subject $S$ is (noninferentially) aware that $S$ is in (Rosenthal 1986, Lycan 2001). By contrast, one is obviously not (immediately) aware of one’s unconscious mental states. The difference lies mainly in how to cash out the expression “aware that one is in.” EHOT theory says that such awareness of $M$ is a distinct (unconscious) state $M^*$ (or HOT) directed at $M$. PSR says that $M^* = M$; that is, $M$ is literally directed back at itself. The WIV says that $M^*$ (i.e. the MET) is an unconscious part of a complex conscious mental state (CMS) directed at $M$ (which is also part of CMS). In each case, some notion of self-reference is involved. This is perhaps more clear for PSR and WIV, but even EHOT theory says that what makes $M$ conscious is a kind of self-referential (unconscious) thought; namely, ‘that I am in $M$’ (see figure X.1).
2. The Wide Intrinsicality View vs. Extrinsic HOT Theory

The main goals of this chapter are to defend the WIV and to argue against PSR. However, it is first very important to mention a few reasons to favor the WIV over EHOT theory. I had originally offered five (in Gennaro 1996: 26-30), but will not rehearse them all here.

First, and perhaps most relevant, is simply that consciousness seems to be an intrinsic property of conscious states. As even Rosenthal has acknowledged (1986: 331, 345), it is preferable to have a theory that can account for this intuitive fact if at all possible. When we are in a conscious state, consciousness does not seem to be analogous to ‘being the cousin of’ or ‘being to the left of,’ but instead seems to be part of the state itself. Of course, most would agree that the reality of conscious states need not match the first-person appearance, and so Rosenthal

<table>
<thead>
<tr>
<th>Introspection</th>
<th>(Extrinsic) Higher-Order Thought Theory (EHOT)</th>
<th>Wide Intrinsicality View (WIV)</th>
<th>Pure Self-Referentialism (PSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconscious HOT (M^{**})</td>
<td>Unconscious MET (M^{**})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Conscious HOT (M^{*})</td>
<td>Conscious MET (M^{*})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mental State (M)</td>
<td>Mental State (M)</td>
<td></td>
</tr>
<tr>
<td>World-Directed Conscious Mental States</td>
<td>Unconscious HOT (M^{*})</td>
<td>Unconscious MET (M^{*})</td>
<td>Conscious State (M)</td>
</tr>
<tr>
<td></td>
<td>Conscious Mental State (M)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure X.1
explains that EHOT theory can still accommodate the phenomenological facts by noting that we are rarely conscious of the HOT itself which renders M conscious. Thus, Rosenthal has recently said that we should not assume “...consciousness reveals everything about our mental functioning, or at least everything relevant to the issue at hand....But to save these phenomena, we need only explain why things appear to consciousness as they do; we need not also suppose that these appearances are always accurate.” (2004: 31) Fair enough, but the real problem is that Rosenthal never presents a compelling case to reject the intrinsicality of consciousness in the first place. For example, he mistakenly argues that if we treat consciousness as an intrinsic property of mental states, then conscious mental states will be simple and unanalyzable (Gennaro 1996: 21-24). Rosenthal defines an “intrinsic” (as opposed to “extrinsic”) property as follows: “A property is intrinsic if something’s having it does not consist, even in part, in that thing’s bearing some relation to something else.” (Rosenthal 1997: 736) But even if consciousness is an intrinsic property of some or all mental states, it surely does not follow that those states are simple or unanalyzable. Conscious mental states can, for example, have the kind of complex structure described by the WIV. Rosenthal sets up a false dilemma: either accept the Cartesian view that mental states are essentially and intrinsically conscious (and so unanalyzable) or accept his version of the HOT theory whereby consciousness, or the so-called “conscious making property” (i.e. being the object of a HOT), is an extrinsic property of mental states. But there is clearly an informative third alternative whereby the MET is part of the overall structure of a conscious mental state. On the WIV, then, consciousness is intrinsic to conscious states, but there is also a kind of inner self-referential and relational element within the structure of such states.

Furthermore, it is not even clear that an intrinsic theorist must rely on any such phenomenological or “intuitive” evidence to make the point. Examining the issue from a third-person, neurophysiological perspective, there is still something odd in holding that what makes a mental state M conscious is something else. For example, if and when the neural correlates of consciousness (NCC’s) are discovered, it seems to me far more likely that such NCC’s will be counted as part of the (global) conscious brain state. That is, what makes a mental state
conscious will be some distributed property of the state itself. There can still of course be a self-referential structure to that conscious state (as the WIV predicts), but both M and MET will be parts of the overall state.

Rosenthal (2004: 33) also rightly demands that an “intrinsic theory must explain what happens when a state goes from being nonintrospectively conscious to being introspectively conscious.” But I have already presented and motivated the WIV version of introspection in both Gennaro 1996 and 2002. Let us first recall that the WIV says that first-order conscious mental states are complex states containing both a world-directed mental state-part M and an unconscious meta-psychological thought (MET). This alternative holds that consciousness is an intrinsic property of conscious states while also providing an analysis of state consciousness. My conscious perception of the tree is accompanied by a MET within the very same complex conscious state. Now when I introspect my perception, there is a first-order mental state which is rendered conscious by a complex higher-order state. Thus, introspection involves two states: a lower-order noncomplex mental state which is the object of a higher-order conscious complex state (see figure X.1 again). I am consciously aware of my mental state. In this case, much like on EHOT theory, the MET itself becomes conscious and is directed at a lower-level mental state.

A second reason to favor the WIV over EHOT theory is that a number of authors have recognized that intrinsic theory seems better suited to avoid several standard problems facing EHOT theory. For example, EHOT theory arguably has a serious problem dealing with the possibility of misrepresentation between the HOT and its target state M (Byrne 1997, Neander 1998, Levine 2001). If we are dealing with a representational relation between two distinct states, it is possible for misrepresentation to occur. If it does, then what explanation can be offered by the EHOT theorist (or HOP theorist for that matter)? If my HOT registers a thought about something green, but M registers a red percept, then what happens? What kind of experience do I have: a reddish or greenish one (or neither)? On either form of intrinsic theory (i.e. PSR or WIV), it seems more difficult to make sense of the possibility of misrepresentation, since either M is directed back at itself (Caston 2002, Kriegel 2003a) or M* is part of the same
state as M (Gennaro 2004).

Another well-known difficulty for EHOT theory has been called “the problem of the rock” (Stubenberg 1998) and “the generality problem” (Van Gulick 2000, 2004), but is perhaps originally due to Alvin Goldman (Goldman 1993). When I have a thought about a rock, it does not thereby make the rock conscious. So why should we suppose, that when one thinks about a mental state M, it becomes conscious? This problem is again based on the notion that there is a complete separation of M* and M. Alternatively, if this aspect of HO theory is rejected and M* is therefore intrinsic to the conscious state in question, then its target must be a mental state “in the head” in the first place, and so such objects as stones cannot be potentially conscious, as the objection suggests (Van Gulick 2000, 2004; Gennaro, unpublished). ⁹

Having said all this, it is necessary to digress for a moment to be very clear about the relative logical space that the WIV occupies, especially given the potential for terminological confusion. First, it will be noticed that the WIV is similar in structure to Rosenthal’s EHOT theory. I do indeed think that there is something importantly correct about HOT theory, and my view admittedly owes much to Rosenthal’s HOT theory. It may even seem that there is no real ontological difference between the two views, but this is not correct for reasons that will be made clearer in section 4. Second, some of the same objections to EHOT theory might also be raised against the WIV, so I have responded to them in print. For example, I have argued at length against the view that HOT theory entails a lack of animal consciousness (Gennaro 1993, 1996, 2004) and against other weak attempts to criticize EHOT theory (Gennaro 2003). Thus, in these cases, I have no problem associating myself with some form of HOT theory and sometimes willingly bring such a characterization upon myself. Of course, if one defines HOT theory as maintaining that the HOT is a distinct state from its target M, then it would not be correct to say that I hold a HOT theory. I do not wish to define HOT theory as such in this way, but others clearly do. In my view, this is more of a terminological dispute, though some seem to have very strong views on the matter. ¹⁰ On the other hand, I have also been urged by some to reject HO theory altogether and endorse a “first-order” (FO) theory, perhaps more along the lines of
Dretske (1995) or Tye (1995, 2000). However, for reasons that go well beyond the scope of this paper, I am not inclined to do so and, in fact, reject FO theory. In any case, I take such accusations from both sides as evidence for the fact that the WIV is a truly viable “middle” position between PSR (and any FO theory) and EHOT theory. Indeed, that is precisely why I originally chose to introduce an entirely new theory name into the literature. It is also why I often use the more neutral expression “meta-psychological thought” instead of “higher-order thought”: there is still a meta-psychological thought about M on the WIV.

3. Against Pure Self-Referentialism
Let us recall pure self-referentialism:

**PSR:** A mental state M of a subject S is conscious if and only if S has a mental state M* which is an appropriate representation of M, and M = M*.

There are a number of somewhat interrelated reasons to reject PSR, which, in turn, will also serve as evidence for the superiority of the WIV.

3.1. *What makes a mental state a conscious mental state?* It is fair to say that all three views do try to answer the question: What is the structure of a conscious mental state? However, a real deficiency for PSR is that it does not even attempt to answer the crucial question: What *makes* a mental state a conscious mental state? Both the WIV and EHOT theory are, in part, trying to explain how an unconscious mental state becomes a conscious one. Of course, many are not satisfied by the explanation offered, but my point here is that PSR does not even attempt to offer any explanation at all. Two reasons for this may be that defenders of PSR are often inclined to reject reductive explanations of consciousness (e.g. Smith 2004: chapter 6; cf. Thomasson 2000: 206) and even to reject the existence of unconscious mental states (Brentano). In this chapter, I will neither argue for the existence of unconscious mental states nor for the view that reductive explanations are desirable. But to the extent that one agrees with one or both of these
assumptions, it seems to me that PSR has a serious problem compared to its rivals. While both the WIV and EHOT theory answer the above question with something like “M becomes conscious when an appropriate HOT (or MET) is directed at M,” PSR can offer no such explanation. PSR does provide a description of the structure of conscious states, but we must distinguish that from the kind of explanation we are seeking. If we ask “What makes M conscious?” for PSR, the response cannot be that M* is directed at M because M is supposed to be identical with M*. How can M* make M conscious or explain M’s being conscious if M* = M? Moreover, either M* is itself conscious or it is not, and then the familiar threat of regress (and even circularity) rears its ugly head. If M* is itself conscious, then what makes it conscious, and so on? Is the consciousness of M being explained in terms of a conscious M*? Also, if M* is conscious, then a reductionist account of state consciousness seems out of the question.

Alternatively, if M* is not conscious, then the PSR defender would, first, have to acknowledge the existence of unconscious mental states, but, even worse, how could M be conscious and M* be unconscious if M = M*?

As I will argue in section 3.5, some supporters of PSR may really hold something more like the WIV. Thus, if we also reject EHOT theory, then the WIV represents a superior middle position. It seems necessary to bring in the notion of parts of conscious mental states in order to give an adequate account of state consciousness. In any case, a fundamental question that should be answered by any viable theory of consciousness goes unanswered or is ignored by PSR advocates. To the extent that M* is introduced merely to articulate the structure of a conscious state, PSR may offer a plausible alternative (though obviously I think it is the wrong structure). But, to the extent that we want M* to play some role in explaining how an otherwise unconscious M becomes conscious, PSR is entirely unhelpful. There is a difference between simply stating that “all (conscious) mental states have a primary and secondary object...,” and giving an explanation for what makes that mental state conscious.  

It might be replied by the PSR theorist that there is a perfectly good explanation for what makes a mental state M conscious; namely, that M becomes conscious when it acquires a
particular sort of self-referential content. That is, M (= M*) has the property of being conscious in virtue of M (= M*) having the property of being self-referential. However, there are at least two problems with this response, relative to WIV and EHOT theory. First, the reply still does not help to provide a reductive explanation of state consciousness (which I am assuming is desirable) because PSR typically holds that one is conscious (in some sense) of the self-referential content itself, as we shall see more clearly below. Recall that I began this chapter by pointing out that all three positions under consideration try to understand state consciousness in terms of some kind of *self-referential intentional* content. Of these three theories, then, it is clear that only PSR cannot offer a reductive explanation in mentalistic terms. Second, I would again insist that the above explanation is not really *explanatory*, but rather *descriptive*. According to PSR, the self-referential content in question is a property of M itself. PSR describes the difference between unconscious and conscious states, but M’s being conscious is not explained by appealing to M* because M* is identical with M. Now perhaps there is some *other* way for PSR to provide a plausible reductive or, at least, naturalistic explanation of state consciousness, but I remain skeptical based on Kriegel’s (2003b: 483, 493-496) “argument from physical implausibility.” For example, suppose we understand a naturalistic explanation to include some kind of *causal* explanation. As Kriegel rightly points out, the causal relation is anti-reflexive and so we can, at most, make sense of such a relation by invoking talk of a part of a complex state directed at another part. The WIV can clearly accommodate the notion of a causal relation between M and M* (in either direction) combining to produce CMS. Notice, however, that there is still no *pure* self-reference; that is, no conscious (brain) state (or state-part) is literally causally related to itself. PSR is indeed physically implausible; it is ruled out if we take the requisite self-reflexiveness literally.

3.2. *Conscious attention cannot be focused both outward and inward at the same time.* Another serious problem with PSR is the failure to recognize the implications of the fact that our conscious attention is *either* world-directed *or* inner directed (but never both at the same time),
as even Brentano acknowledged (1874/1973: 128-129). When I am assembling a bookcase or working on this paper, my conscious attention is focused outside of me, e.g. at the bookcase or at my computer screen. Of course, if either the WIV or EHOT theory is correct, what makes that state conscious is an unconscious HOT (or MET) directed at M. If, however, I reflect or introspect on my experience, then my conscious attention is focused inward at the mental state itself. But PSR cannot provide such a neat explanation of the difference between outer and inner directed consciousness. Leaving aside regress worries, if M* is supposed to be conscious and directed back at the entire conscious mental state M, then it would seem that M* is both directed at the world and at one’s own mental state M at the same time because, after all, M is supposed to be identical with M*. This doesn’t seem possible, if it is even coherent at all. I certainly may frequently shift my attention between, say, the bookcase and my experience of working on it, but I never consciously focus on both at the same time. It therefore also seems that proponents of PSR often slide back and forth between outer-directed consciousness and introspective consciousness without even realizing it. But if PSR is to give an account of a world-directed conscious state, M, then it seems committed to the absurdity that M is both directed at the world and at itself at the same time. Of course, if M* is not itself conscious (as the WIV and EHOT views have it), then PSR either has the problem mentioned in section 3.1 (i.e. how can M = M* if M is conscious and M* is unconscious?), or we end up with something closer to the other theories. In any case, PSR supporters do believe that M* is itself conscious in some sense, as we shall see. Unlike the WIV, there is no explicit belief in (unconscious) “parts” of conscious mental states. But the WIV has the advantage of holding that M is an outer-directed conscious part of a complex conscious state (CMS) within which an (unconscious) M* is directed at M. Bringing in parts of conscious states seems unavoidable if one wants to preserve some kind of self-reference in state consciousness.

Somewhat related objections to PSR are raised by Perrett (2003) and Zahavi (1998). For example, Perrett argues that “...there is an inconsistency in Brentano’s account. On the one hand, he holds that the content of an awareness is always a proper part of that awareness, where a
**proper part** is a part that is not identical with the whole of which it is a part. On the other hand, the secondary awareness is also supposed to possess a content which is identical with itself, since it is its own object. Thus the content of such an awareness cannot be a proper part of itself” (2003: 231). As I argued above, then, if \( M = M^* \), then \( M^* \) would have to be directed back at \( M \) *in its entirety*; that is, directed back at \( M \) and \( M^* \). But if \( M^* \) is itself conscious, then \( M^* \) is *both* directed at the world *and* at one’s own mental state \( M \) at the same time. Similarly, Zahavi (1998: 139) notices a “disastrous problem” using Brentano’s example of hearing a sound or tone: “A [conscious] act which has a tone as its primary object is to be conscious by having itself as its secondary object. But if the latter is really to result in self-awareness, it has to comprise the entire act, and not only the part of it which is conscious of the tone. That is, the secondary object of the perception should not merely be the perception of the tone, but the perception which is aware of both the tone and itself.” The WIV has no such problems. An unconscious MET is directed at only *part* of the entire CMS; i.e., the \( M \) which is consciously directed at the world. Moreover, the MET is therefore not consciously directed at \( M \), avoiding the conflation with introspection.

Now, it might be objected at this point that I have ignored a crucial distinction, frequently made by PSR advocates, between *attentive* (or “focal”) consciousness and *inattentive* (or “marginal” or “peripheral”) consciousness. Not all conscious “directedness” is attentive, and so perhaps I have mistakenly restricted conscious directedness to that which we are consciously focused on. The idea is that, in figure X.1, the “back-turning” arrow for PSR represents inattentive (inner-directed) consciousness whereas the other arrow represents focused (outer-directed) awareness. If this is right, then my objection has an easy reply; namely, that first-order conscious state \( M \) is both attentively outer-directed and inattentively inner-directed. \( M^* \) is thus conscious in this inattentive sense. I have three replies to this for now. First, although it is probably true that there are degrees of conscious attention, it seems to me that the clearest examples of “inattentive” consciousness are still outer-directed; for example, perhaps some of the awareness in one’s peripheral visual field while watching a concert or reading a book. But
this obviously does not show that any such inattentional consciousness is *self-directed* at the same time when there is outer-directed attentional consciousness. Second, what is the evidence for such self-directed inattentional consciousness? It is based on phenomenological considerations. Now, I will tackle the phenomenological argument more directly and thoroughly in section 3.4 but, to anticipate, it suffices to say for now that I do not find such inattentive “consciousness” in my experience, which should presumably show up in the Nagelian sense if it is based on phenomenological observation. Conscious experience is often so clearly and completely outer-directed that I deny we have such marginal self-directed conscious experience when in first-order conscious states. It does not seem to me that I am consciously aware (in any sense) of my own experience when I am, say, consciously attending to a movie or the task of building a bookcase. As we shall see, this point has greater force when we consider the conscious experience of animals or infants; it seems even more unlikely that they are capable of having such peripheral self-directed consciousness. Third, when PSR theorists claim to find such inattentive consciousness in their experience, a case can be made that they are philosophically “reflecting” on their experience, but then they are *consciously attending* to their experiences, which is really introspective consciousness (as we have already seen). Thus, we no longer have a phenomenological analysis of first-order conscious states.

3.3. *PSR offers no plausible account of introspection.* It is also curious that no clear account of introspection has been presented by supporters of PSR. Perhaps this is simply due to the fact that they are mainly concerned with first-order conscious states. However, the problem goes much deeper than this for several reasons. First, as we saw above, PSR theorists often conflate introspective consciousness with an explanation of first-order conscious states. If M* is itself conscious, then that seems to indicate the presence of an introspective state, not merely a first-order conscious state. Second, some theorists who otherwise oppose any form of HOT theory are sympathetic to it as an account of introspection; that is, when one is *consciously aware* of a mental state (Block 1995). Rosenthal then reasons as follows: “...if a state isn’t conscious [at all],
there is no HOT. That suggests that a state’s being conscious in the ...[world-directed]
nonintrospective way results from something in between these two [i.e. a nonconscious HOT or
MET].” (Rosenthal 2004: 24) Both EHOT theory and the WIV can accommodate this important
aspect of a theory of consciousness. On the other hand, PSR offers no explanation of just how
such a transition from first-order conscious states to introspective states might occur. Third, if I
am right thus far, it is simply difficult to understand what the structure of introspective
consciousness would be according to PSR. If no unconscious thought becomes conscious during
such a transition, then does an entirely new state, M**, emerge as directed at M (and therefore
also at M*)? Is M** itself also conscious (on pain of regress)? Would M** also be “directed
back at itself” so we would then also have an M*** directed at M**?18

3.4. The phenomenological argument. Perhaps most importantly, PSR supporters argue that M*
is conscious (in some sense) based on phenomenological considerations (Kriegel 2003a, Smith
1986). Of course, in Brentano’s case, M* would have to be conscious because he did not believe
in unconscious mental states. There are, as we have already seen, significant problems with
holding that M* is conscious, but I now wish to challenge this view more directly as it is likely at
the root of the above difficulties.

Focusing on Kriegel’s 2003a paper (cf. 2003b: 485), we find a distinction between
“intransitive self-consciousness (or self-awareness)” and “transitive self-consciousness.” He first
rightly explains the latter in much the same way that EHOT theory and the WIV speak of
introspection: “...a transitive self-conscious state is introspective, in that the object is always
one of the subject’s own mental states....” (2003a: 105) On the other hand, “...an intransitively
self-conscious state is ordinarily not introspective, in that usually its object is an external state of
affairs.” So far so good. Like the WIV, each first-order conscious state contains a meta-
psychological component (which is indeed a form of self-consciousness or self-awareness on my
view), but the conscious state is outer-directed. Moreover, when the shift to introspection occurs,
then there is a conscious meta-psychological thought directed at one’s own mental state. As
Kriegel also makes clear (on p. 105), in such transitively self-conscious cases, there is a further intransitive self-consciousness accompanying that conscious state.

The key point here lies in the fact that Kriegel takes intransitive self-consciousness (i.e. M*) itself to be conscious based on phenomenological observation. As was mentioned in section 3.2 in addressing “inattentional” consciousness, I strongly disagree and so am much closer to EHOT theory, at least in this respect. M* is not conscious in any meaningful sense of the term, including the Nagelian sense. For one thing, Kriegel uses a number of vague, but typical, characterizations of M*, e.g. “subtle awareness of having M,” “implicit awareness of M,” “dim self-awareness...humming in the background of our stream of consciousness,” and “minimal self-awareness” (2003a: 104-5). Now I have no objection to these expressions as such, but it is still not at all clear to me that M* is conscious in any phenomenological sense. Rather, I think it is far better to construe M* as unconscious, and so I often speak of unconscious “meta-psychological thought awareness” and “non-reflective self-consciousness” in Gennaro 1996 and 2002. Much like we normally understand, say, at least some peripheral vision to be unconscious, the same should go for the “subtle awareness” in question. It also does not help to speak of M* as “experienced” or as an “experiential state” (Kriegel 2003a: 121ff.) for this begs the question as to whether or not this “awareness” is phenomenologically consciously experienced. Nor does it help to appeal to “inattentional” consciousness.

Three more specific problems come to mind. First, the examples used by Kriegel to illustrate the consciousness of M* really cause us to shift (phenomenologically) to transitive consciousness (i.e. introspection). We are asked (on p. 104), for example, to “suppose...that you suddenly hear a distant bagpipe. In your auditory experience of the bagpipe you are aware primarily, or explicitly, of the bagpipe sound; but you are also implicitly aware that this auditory experience of the bagpipe is your experience...” But it seems to me that the very act of performing this mental exercise results in an act of introspection or reflection. That is, Kriegel is asking us, via our imagination, to focus consciously on M (e.g. the experience of hearing a distant bagpipe) in “considering” his examples. We are really asked to reflect on the hypothetical
case in question. How can we pretend to “consider” such a state of mind without shifting our phenomenological attention onto the mental state or experience itself? To the extent that we really can do so, for what it’s worth, I think that our consciousness is completely outer directed, for example, when I am absorbed in a taxing chore or taken with a beautiful painting. Such conscious states can still have the structure of the WIV, but there is no conscious notice at all of \( M^* \) (= MET). In a sense, then, although Kriegel is not fallaciously conflating the distinction between introspection and first-order conscious states, he is, I believe, relying on one’s reflective response to his examples in order to make the case that \( M^* \) is conscious. It is crucial to remember that the WIV also holds there is an implicit self-referential MET as part of the overall conscious state, but this is not to say that the MET is itself conscious. Indeed, if it were, then we would have a case of introspection, not a world-directed conscious state.

Second, if Kriegel’s (or any) phenomenological argument is meant to support PSR, then it also fails due to the lessons learned from sections 3.1 and 3.2. If we cannot simultaneously \textit{consciously} attend to both outer objects and inner mental states and \( M = M^* \), then how could \( M^* \) be conscious? The very same state (without any “parts”) would then be both outer-directed and inner-directed, which is impossible. Moreover, if \( M = M^* \) and \( M^* \) is itself conscious, it is difficult to understand how the presence of \( M^* \) can help to explain why \( M \) is conscious in the first place, especially in any reductivist sense.

Now, Kriegel is right that “there is something artificial in calling a mental state conscious when the subject is \textit{wholly} unaware of its occurrence” \( (2003a: 106, \) emphasis added). But this leaves open whether such awareness is conscious or not. Kriegel’s use of the expression “wholly unaware” suggests both “consciously and unconsciously unaware,” but we might instead hold that a state is conscious when the subject is \textit{unconsciously} aware of its occurrence. Kriegel’s argument can really only justify the weaker claim that “there is something artificial in calling a mental state conscious when the subject is \textit{not at least unconsciously aware} of its occurrence.” But this is precisely one key issue at hand between PSR and the WIV.\(^{19}\)

Third, the above is particularly important to the extent that we want a \textit{general} theory of
state consciousness. That is, although it is certainly true that there are degrees of conscious attention and degrees of self-consciousness, it is desirable to offer an explanation of what all first-order conscious states have in common. On the WIV, this is the fact that there is an unconscious MET directed at M, both of which are parts of a complex CMS. I suggest that such a general account can only be offered if the MET (or M*) is itself unconscious because we are so often entirely consciously focused on outer things. Moreover, if we are to allow for, say, animal and infant consciousness, the notion that M* is itself conscious seems very unlikely. Indeed, believing otherwise could lead to the problematic conclusion that such creatures are incapable of having conscious states at all. Instead, contra Kriegel, it seems better to hold that any genuine case of a conscious M* is really an instance of introspection, and thus any first-order conscious state is only accompanied by an unconscious M*. In short, the problem of animal and infant consciousness applies with even greater force to PSR than to either WIV or EHOT theory.

The importance of all this can also be seen in Smith’s (2004, Chapter 3) most recent account where he retreats from what appeared to be a previous adherence to PSR (in Smith 1986, 1989). Smith continues to insist that “...the formal analysis of inner awareness [M*]...is a task for phenomenology.” (2004: 80) However, this leads him to abandon his earlier view that all first-order conscious states have such inner awareness (Smith 2004: 109-116). Smith now allows for basic levels of outer-directed consciousness which lack such self-consciousness or inner-awareness; for example, when “I am unselfconsciously hammering a nail, or driving down the highway, or choosing to hit the tennis ball crosscourt rather than down the line....” (2004: 109) Thus, “On the view now emerging, inner awareness is an integral part of higher levels of consciousness, realized in humans and perhaps other animals, but it is not present in lower levels of consciousness in humans and other animals.” (2004: 110) To the extent that Smith no longer advocates PSR, then I agree. However, I suggest that what he should really give up is the view that such “inner awareness” [= M*] is phenomenologically revealed. If he had done so, then he would have recognized that all outer-directed conscious mental states can still have a WIV-like structure without giving up the belief that inner-awareness or self-consciousness (of some kind)
is indeed built into the structure of those states. Smith is correct, however, in recognizing the existence of conscious mental states not accompanied by a conscious M*. But this should lead one to embrace something more like the WIV instead of the belief that there can be levels of consciousness without any inner awareness whatsoever, especially if one is sympathetic to any of the three positions under consideration.

Thus, I disagree with Kriegel that the only reason to posit “...such [intransitive] self-awareness...is on first-personal experiential grounds” (2003a: 121) and it also not true that “those who insist that they do not find in their experience anything like an awareness of their conscious perceptions and thoughts probably deny the very existence of intransitive self-consciousness.” (2003a: 121) In the absence of such alleged phenomenological evidence, it is quite appropriate (as Kriegel does) to demand other theoretical and explanatory advantages to positing such self-awareness. Although by no means conclusive, it seems to me that there is ample reason to posit such unconscious METs.²⁰

In some ways, then, the phenomenological argument is at the root of the problems raised for PSR in sections 3.1-3.3. If one believes that M* is conscious, then (a) it is difficult to offer any reductionist explanation of state consciousness, (b) one is more likely to conflate introspection with first-order conscious states, and (c) one is unable to offer an account of introspection.

3.5. Many PSR views are really the WIV in disguise. Finally, it seems to me that, to the extent that PSR is plausible at all, it is really better construed as the WIV anyway. A close reading of the literature reveals at least some evidence for this claim. Talk of “parts” and “wholes” of conscious states abound even when characterizing Brentano’s views, not to mention Kriegel’s own very useful analysis (in this volume). It is doubtful that all such references can easily be explained away as merely metaphorical.

For example, consider the following sampling of quotations (emphases added). “The presentation which accompanies a mental act and refers to it is part of the object on which it is
directed” (Brentano 1874/1973: 128). This suggests that, even for Brentano, M* [= “the presentation...”] is really only part of the “entire” conscious state. In describing Brentano’s view, Natsoulas similarly explains: “Not only is a conscious mental-occurrence instance presented and directly apprehended..., but also there is, as part of its occurrence, awareness of it as this (a mental-occurrence instance).” (Natsoulas 1993a: 117) And Smith once said that inner awareness “must be an occurrent part of the given mental event itself.” (Smith 1989: 81) It is difficult to see how M could be identical with M* when reading such passages. We are also told by Smith that Brentano’s secondary inner consciousness is “a dependent, inseparable part of the given act [of consciousness].” (Smith 2004: 78-9; cf. 93; see also Thomasson 2000: 192, 196)

Caston (2002: 792-793) also discusses the importance of the part-whole relationship in his analysis of Brentano and Aristotle. Perhaps most interesting is Caston’s diagram (2002: 778) presumably representing PSR. But much like the WIV for outer-directed conscious states, we have an arrow going from one part of a complex conscious state (the “perceiving”) to another part (the “seeing”) divided by a broken line, in addition to the arrow representing the outer-directedness of the entire conscious state. Caston then speaks of both aspects as essential to any token perception.21

There has also been a noticeable emphasis on parts and wholes in the more recent writings of Kriegel, who, as we have seen, did argue for PSR in Kriegel 2003a.22 In addition to Kriegel’s chapter in this volume, he had previously said that “...a brain state can be said to represent itself if one part of it represents another part of it.” (2003b: 493; italics in original) This comes in the context of preserving some kind of self-representational view within a naturalistic framework (as was noted in section 3.1), but it also sounds much more like the WIV than PSR. Moving more clearly away from PSR, Kriegel also says that “the mental state in question does not actually represent itself. At most, we can say that one part of it represents another part.” (2004: xxx)

The case against PSR is very strong. If we are to preserve any useful notion of self-reference at all, something more like the WIV is necessary.
4. Objections and Replies

For the sake of further defending and clarifying the WIV, let us now consider a number of objections.

4.1. The unconscious parts objection. It might first be objected that the notion of an unconscious part of a complex conscious state is difficult to understand. What does it even mean to say that state-parts can be conscious or unconscious? The WIV is an implausible view perhaps because the notion of an unconscious part of a conscious state sounds contradictory.23

There are at least two replies to this. First, the objection could be taken as committing the well-known “fallacy of division”; namely, that what is true of the whole must be true of each part of the whole. Water extinguishes fire, but oxygen does not. If this is the motivation behind the objection, then it clearly fails. Indeed, if we think of it in a physicalistic way, then such logic would seem to lead automatically to panpsychism. Why should we suppose that each part of a complex conscious state is itself conscious? I see no reason at all to assume that one will be consciously aware of each intrinsic part of a conscious state. Are we consciously aware of everything involved in a conscious state? I suggest not, and this is particularly clear if we think about conscious states as globally represented brain states. All kinds of nonconscious mental activity are involved in a conscious state. Although I disagree with Colin McGinn’s “mysterian” views on consciousness, he is right when he speaks of there being a “hidden structure of consciousness” such that there are “surface properties, which are accessible to the subject introspectively; and deep properties, which are not so accessible.” (McGinn 1991: 111) Indeed, on the WIV, the MET is precisely such a deep property, but can surely still be part of (i.e. intrinsic to) the entire complex conscious state. As McGinn puts it: “...the subject is not conscious of the deeper layer...but it does not follow that this layer does not belong intrinsically to the conscious state itself. Just as F can be an intrinsic property of a perceptible object x without being a perceptible property of x, so conscious states can have intrinsic properties that they do not have consciously.” (McGinn 1991: 98).24
Second, we must distinguish between the first-person and third-person perspectives on conscious states. From the first person point of view, we cannot expect to be consciously aware of all that is “presupposed,” to use a Kantian term, in a conscious state. As I have argued at greater length (in Gennaro 1996, Chapter 3), we can understand the situation as follows: We passively receive information via our senses in what Kant calls our “faculty of sensibility,” some of which rises to the level of unconscious mental states. But such mental states do not become conscious until the more cognitive “faculty of understanding” operates on them via the application of higher-order concepts in the METs. Thus, I consciously experience the brown tree as a brown tree partly because I have the MET “I am seeing a brown tree now.” Once again, however, the MET itself is not conscious; thus, METs and their concepts are “presupposed” in conscious experience. The understanding unconsciously “synthesizes” the raw data of experience in order to produce the resulting conscious state. As Kant understood very well, there must be significant unconscious (synthesizing) activity implicit in each conscious state, and we are not conscious of that activity itself despite the fact that it is intrinsic and essential to the resulting conscious state. Indeed, it is the MET that makes the state conscious because of the conceptual activity directed at the lower-order mental state. As most would also agree, concepts are constituents of thoughts, which I construe as momentary episodic mental events unlike dispositional states (such as beliefs).

There is nothing at all implausible about the existence of unconscious parts of conscious states. On the contrary, such a view is crucial to appreciating the subtlety of the WIV. 25

4.2. Rosenthal’s objection. David Rosenthal (1993) objects that HOTs (or METs) cannot be intrinsic to conscious states because it would be contradictory for a single state to be, for example, a conscious doubt that it is raining and an affirmative (i.e. assertoric) thought that I am in that state. That is, we normally individuate states in terms of mental attitude.

In order to reply, let us make an important threefold distinction. First, we have the conscious state, that is, the vehicle which is identical with a mental representation and is
presumably a brain state of some kind. Second, there is the representational content of the state in question; that is, what the state is about or directed at. Third, there is the mental attitude (or “mode”) of the state; that is, what type of mental state it is, e.g. a doubt, a thought, a perception, and so on. This threefold distinction is particularly crucial when teasing apart the sometimes subtle differences between the three views under consideration (see e.g. also Hossack 2003: 192, 200; Kriegel 2003b; 2004; Smith 2004, Chapter 3). Now, unlike EHOT theory, virtually any form of intrinsic theory will hold that a single vehicle or state can have dual representational content. More specifically, the WIV (and PSR for that matter) says that the higher-order content is represented in the same state as the first-order content. This view seems perfectly defensible, though I will not argue for it here. More to the point raised by Rosenthal, it also seems possible to have a single conscious state involving two attitudes because one will be directed at its first-order content and the other will be directed toward its higher-order content. Thus, on the WIV, a complex conscious state, CMS, can have one attitude (e.g. a doubt) directed at the weather and another (an assertoric thought) directed at the doubt. M and MET can be instances of two different attitudes and yet nonetheless be parts of a single vehicle or brain state. As Kriegel (2003b: 486-488) persuasively argues, a subject’s conscious state containing two attitudes would still be related differently to its two contents and a case has already been made for this view in some areas of moral psychology. There is nothing contradictory here.

Unlike PSR, however, on the WIV M* (= MET) is not directed back at the entire CMS, as we saw in section 3.2. To reinforce the point, then, we can now symbolize WIV as follows:

\[ S \{A1 (C1) \& A2 (C2)\} \]

where S stands for a (conscious) state which includes everything in the square brackets, i.e. two attitudes (A1 & A2) along with their respective contents (C1 & C2). But what is crucial is that A1 is part of C2, that is, the second inner-directed attitude (A2) is directed at A1. On the other hand, even if PSR has the same formal structure, one problem results from the notion that somehow C2 = S. This is an untenable view, as was argued in section 3. In contrast to both, EHOT theory says: S1 \{A1 (C1)\} and S2 \{A2 (C2)\}, where S1 and S2 are distinct states.
4.3. A “sum” or “complex” account? This leads to another closely related question: What exactly is the real ontological difference between the WIV and EHOT theory? Some have claimed that the difference is merely terminological or verbal. Indeed, Rosenthal himself once said that there is no non-arbitrary way of choosing between these two ways of describing higher-order theory (1986: 345).

There really are two issues here. First, Kriegel has convincingly argued that whether there are two states or one state is not at all arbitrary and can even be understood as an empirical neurophysiological matter (Kriegel 2003b: 488-494). Citing the familiar “binding problem,” he explains how the difference may simply depend on whether or not two neural events, N1 and N2, taking place in different parts of the brain “...are synchronized or not. If they are, then N1 and N2 are bound into a single brain state; if they are not, N1 and N2 constitute two separate brain states....” (Kriegel: 2003b: 493) I see no reason why a proponent of the WIV cannot also agree, at least as a partial explanation of the real ontological difference between the WIV and EHOT theory.

Second, there is the issue of just what the nature of such “compound” states are. Kriegel (this volume) usefully distinguishes between what he calls “sum” and “complex” theories. This division aligns with his earlier (2004) distinction between the mere compresence of two mental states and the integration of mental states. The basic difference is that a mere sum theory says that what makes two states part of a single mental state is merely our decision to treat them as such. Once again, it is a purely verbal or stipulative difference which Kriegel (2004) calls the “conceptual-relation strategy.” In contrast, a complex is a sum whose parts are essentially connected, or bound, in a certain way. There is a psychologically real relation or integration between the parts; hence, Kriegel (2004) calls it the “real-relation strategy.”

Now Kriegel (2004, this volume) construes the WIV as a sum account whereas he understands his own “cross-order information integration” (or “same-order monitoring”) model and Van Gulick’s HOGS model as complex accounts. I strongly disagree with this
characterization of the WIV, though it is perhaps understandable how one might take it that way, given the admittedly embryonic form the theory took in my 1996 book. However, I do think there is strong evidence to indicate that the WIV is also a complex account, even from Gennaro 1996. First, in rejecting Rosenthal’s EHOT theory, I spoke of how the “...very nature of conscious states is colored by the concepts [in the METs which are] brought to bear on them.” (Gennaro 1996: 29) I urged that the MET actually changes the nature of the conscious state, so that, unlike EHOT theory, the object of a MET is not merely passively there unaltered by the MET. Second, I criticized Rosenthal’s belief in unconscious qualitative states because the conceptual activity in the METs is essential to the very identity of the overall conscious state it is part of. So “a nonconscious qualitative state, contra Rosenthal, could not be the very same state as the conscious one because of the lack of conceptualization.” (Gennaro 1996: 30) That is, when M becomes conscious as part of a CMS, it is not just the same state with consciousness added to it. Third, as was mentioned in section 4.1, I elaborated on and emphasized (even in neurophysiological terms) the Kantian-style thesis that it takes appropriate interaction and cooperation between the “sensibility” and the “understanding” (i.e. M and MET) in order to produce the resulting conscious state (Gennaro 1996: Chapter 3). More recently, I have briefly discussed the importance of “feedback loops” in the brain in order to illustrate the essential mutual interaction between the parts of a conscious state (Gennaro 2004: 62-63). The WIV, thus, embodies a real-relation strategy and is a complex, not sum, account of state consciousness.

In light of this and section 4.2, it is worth noting that, in a very striking passage, Rosenthal (2002: 416, emphasis added) himself says that “On the HOT hypothesis, a conscious state is a compound state, consisting of the state one is conscious of [i.e. M] together with a HOT. So the causal role a conscious state plays is actually the interaction of the two causal roles...This explains how a state’s being conscious may to some extent matter to its causal role.” This comes in response to Dretske’s (1995: 117) charge that HOT theory is unable to explain how a mental state’s being conscious could have any function; that is, it would seem that the state’s being conscious would make no difference to its causal role. HOT theory thus threatens to
make consciousness merely epiphenomenal. Although Rosenthal is somewhat reluctant to concede the point, he does, in the end, seem to agree that there is at least one important difference between WIV and EHOT theory; namely, that only the former can explain how the causal/functional role of a conscious state can be importantly different from the relevant target state stripped of its HOT. It seems to me he rightly suggests that (at least sometimes) M and M*, when combined, can form a uniquely new state, at least in terms of its functional role, which is therefore not a mere “sum” of M and M*. This sounds more like the WIV than EHOT theory, particularly given Rosenthal’s note that the “…interaction of the two roles may not be [merely] additive…” (Rosenthal 2002: 421, note 48).

4.4. Droege’s objection. We are now in a much better position to respond to Paula Droege’s (2003) complaint that “on Gennaro’s theory, it is unclear whether the whole complex state is conscious and there is something it is like to be in only part of it, or whether there is something it is like to be in the whole complex state....If the whole, then the theory is circular....If part, then either [M] is what is conscious and we are back to [EHOT theory], or the [MET] is conscious and the theory is again circular.” (Droege 2003: 40-1)

This is certainly an important request for clarification. The first part of the answer is that, on my view, it would be most precise to say that “the whole complex state [= CMS] is conscious and there is something it is like to be in only part of it [= M],” but there are important ambiguities here. It is crucial to recall the distinctions between the first-person and third-person perspectives (section 4.1) and between a conscious state (or vehicle) and its content (section 4.2). Looking at it from a more third-person point of view, we can then see that the entire complex (brain) state should be understood as conscious. Of course, there is still something it is like to be in the whole state (CMS) if that merely means that when a subject S is in CMS, S is having a subjective phenomenal experience. However, there is something it is like to be in only part of CMS in the sense that S is only consciously aware of the content of M from the first-person point of view. We can and should deny Droege’s conclusion that we are then “back to EHOT theory”
because of what we have seen in the previous sections (4.2 and 4.3). Since the WIV is a “complex” account, the very conscious content of M is essentially interwoven with the MET. That is, M would not have the content it has without its relation to the MET or if the MET were a completely distinct state from M. The MET is presupposed in the very nature of M’s conscious content and is thus part of the same state as M. Moreover, we have seen many other reasons to distinguish between the WIV and EHOT theory. The WIV is also not circular because it still attempts to reduce state consciousness to the interaction between two unconscious mental states, which are fused together to bring about a unique complex conscious state. This is not a claim that can be made by PSR, as we saw in section 3.  

4.5. Schröder’s objection. Let us also consider the remark by Jürgen Schröder (2001) that it is “doubtful whether [the WIV] does really account for our intuition that consciousness is an intrinsic property of our mental states” mainly because “although consciousness is now intrinsic to the whole state [= CMS], it is not intrinsic to the mental state which is a part of the whole [= M]. This is so because the conscious-making thought [= MET] is not a property of [M], but a mental state of its own.” (2001: 35, note 8) That is, since I have just acknowledged that there is only something it is like to be aware of M’s content, then doesn’t this undermine the initial motivation to make consciousness intrinsic to state consciousness (e.g. early in section 2)?

The answer is no, but we must again be precise. First, since the WIV is a “complex” account, it is not quite right to say that “the conscious-making thought [= MET]...is a state of its own.” Once again, the MET is part of (i.e. one attitude within) the very same state that M is also part of. Second, being represented by the MET is a property of M, at least in the sense that the MET contributes essentially to M’s conscious content. Third, Schröder seems mistakenly to equate “consciousness” with “the conscious-making thought [= MET].” As we have seen, the MET is itself not conscious at all. Finally, then, the first-person intuition that “consciousness seems intrinsic to conscious states” should be understood as a combination of two claims; namely, that consciousness seems intrinsic to M’s content from the subject’s first-person point of
view, and that consciousness is intrinsic to the state that M is part of. When it seems to me from
the first-person point of view that consciousness is not extrinsic to, say, my conscious perception
of a house, I am reflecting on the common sense “intuition” that consciousness seems intrinsic to
my first-person conscious awareness of the house and also to the perceptual state which includes
that conscious awareness. The WIV does accommodate this intuition.33

4.6. The infallibility objection. One final objection to the WIV is the charge it entails that
knowledge of one’s conscious states is infallible, especially in light of the reply to the problem of
misrepresentation noted in section 2. If M and M* cannot really come apart, then doesn’t that
imply some sort of objectionable infallibility?34

This objection once again conflates outer-directed conscious states with allegedly
infallible introspective knowledge. On the WIV, it is possible to separate the higher-order
(complex) conscious state from its target mental state in cases of introspection (see figure X.1
again). This is as it should be and does indeed allow for the possibility of error and
misrepresentation. Thus, for example, I may mistakenly consciously think that I am angry when I
am “really” jealous. The WIV properly accommodates the anti-Cartesian view that one can be
mistaken about what mental state one is in, at least in the sense that when one introspects a
mental state, one may be mistaken about what state one is really in. However, this is very
different from holding that the relationship between M and M* within an outer-directed CMS is
similarly fallible. There is indeed a kind of infallibility between M and M* on the WIV, but this
is not a problem. The impossibility of error in this case is merely within the complex CMS, and
not some kind of certainty that holds between one’s CMS and the outer object. When I have a
conscious perception of a brown tree, I am indeed certain that I am having that perception, i.e. I
am in that state of mind. But this is much less controversial and certainly does not imply the
problematic claim that I am certain that there really is a brown tree outside of me, as standard
cases of hallucination and illusion are meant to show. If the normal causal sequence to having
such a mental state is altered or disturbed, then misrepresentation and error can certainly creep in
between my mind and outer reality. However, even in such cases, it is usually not doubted that I am in that conscious state at all.

5. Conclusion

The WIV is a more plausible account of state consciousness than either PSR or EHOT theory. It is importantly located between its rivals in such a way as to avoid the problems of each while retaining their virtues. We can and should acknowledge that some form of self-reference is involved in any conscious mental state. However, conscious mental states are not literally directed back at themselves; nor are they made conscious by entirely distinct HOTs. Conscious mental states (CMS) are complex states with parts such that an unconscious MET is directed at a world-directed M. Any trend in this direction in the recent consciousness literature should be looked upon as a positive one.35

Notes:

1. Brentano’s own view is actually much more elaborate than this. For example, Brentano (1874/1973: Book. 2, Chapter 3) actually holds that there four different aspects included in every mental act, including a feeling toward itself. However, for my purposes, these very problematic details are irrelevant and can be safely ignored.

2. Following Rosenthal (e.g. 1997), I use the expression ‘state consciousness’ since I am mainly concerned with consciousness as a property of mental states (as opposed to ‘creature consciousness’). By ‘conscious’ or ‘conscious state’ I normally have in mind Nagel’s (1974) “what it is like” sense.

3. See, for example, Rosenthal 1986, 1993, 1997, 2002, 2004. I call his HOT theory “extrinsic” in this chapter only to emphasize the fact that his theory regards the HOT as a distinct state from its target. Rosenthal (2004: 30-35) himself has spoken of “intrinsic higher-order thoughts” in contrast to his view. I will also not explicitly address the so-called “higher-order perception
(HOP)” model (e.g. Lycan 1996), though it is also an extrinsic account and much of what I say in contrast to EHOT theory also applies to HOP theory.

4. I have in mind Uriah Kiegel’s “cross-order information integration” model (2004, this volume; cf. also his 2002 and 2003b) and Robert Van Gulick’s (2000, 2004) “higher-order global states” (HOGS) model, though I hasten to add that I owe a great intellectual debt to them for helping me to think more clearly and deeply about my own view.

5. It should be noted that Jean-Paul Sartre is also often cited as holding PSR. In Gennaro 2002, I argue that he held something closer to the WIV, though I did not explicitly address PSR in that paper. I will not address Sartre’s view here.

6. This is therefore similar to Kriegel’s SOMT2 (this volume), but, as we shall see, I actually hold something closest to his SOMT10. When I say that M and M* are ‘proper parts’ of CMS, I basically mean that they are parts which are not identical with the whole of which they are parts. Thus, for example, M* cannot be part of itself; nor can the CMS be part of M*. It might be objected that my definition is circular, or at least nonreductionist, since the term ‘conscious’ appears on each side of the biconditional. However, I do not think that this is case, as I hope will become clear by the end of the paper (especially in section 4). As we shall see, the WIV is still reductionistic because an unconscious M* (= MET) is what makes an otherwise unconscious M conscious. The definition is also not circular because there is a crucial ambiguity in how the term ‘conscious’ is used. The ‘conscious M’ is meant to refer to the first-person subjective point of view whereas speaking of the ‘whole CMS’ is based more on third-person considerations. I chose to put M, instead of CMS, on the left side of the biconditional in order to more clearly express the differences among the three definitions (unlike Kriegel’s SOMT10 which begins with the “whole” on the left side of the biconditional).

7. This is virtually the same as Kriegel’s SOMT1 (this volume).

8. And, unlike Descartes and Brentano, we certainly need not suppose that all mental states are conscious. I also show that Rosenthal mistakenly argues that intrinsicality entails essentiality and that extrinsicality entails contingency (in Gennaro 1996: 21-24).

10. Kriegel agrees that this is largely a terminological matter, but then still opts to restrict use of ‘higher-order’ to those theories which treat the HOT as a distinct state. Hence, he calls his view “same-order monitoring” (in this volume). However, during a session at the 2004 “Toward a Science of Consciousness Conference” in Tucson, it also became clear that there are stronger views on this matter. Andrew Brook urged me to jettison all use of ‘higher-order’ in my theory whereas Peter Carruthers thought that Kriegel had misnamed his theory. I agree more with Carruthers here, and Van Gulick also clearly has this preference; but, again, I take this mainly to be a terminological dispute. One problem, though, is that the converging similarities between all of these positions might be lost.

11. By Andrew Brook and Robert Lurz, in conversations.

12. For example, Dretske’s theory seems incapable of providing a principled distinction between unconscious and conscious mental states (see e.g. Thomasson 2000: 200-202 for some discussion). See also Carruthers (2000: Chapter 6) and Lurz (2004) for additional reasons to reject standard FO theory.

13. In retrospect, perhaps I should have chosen a more “catchy” name for my theory, but, at this point, I am hesitant to add to the abundance of acronyms and theory names already in the literature. Even just “WIT” (wide intrinsicality theory) would at least have been easier to say. Other sexier possibilities are “intrinsic HOT theory” (IHOT) and the more provocative “1½ order theory of consciousness” or “split-level theory of consciousness.”

14. For what appears to be an intrinsic version of HOP theory, see Lormand (unpublished).

15. For some discussion of Sartre on this matter, see Gennaro 2002: 299-308.

16. David Woodruff Smith (e.g. 2004: Chapter 3) is also a good example of someone concerned solely with the structure (or “form”) of conscious states at the expense of offering any kind of
explanation as such. Of course, I do not claim to have solved any kind of “hard problem” in this chapter, but see Gennaro (unpublished) for one attempt.

17. It may be useful to distinguish between “momentary focused introspection” and “deliberate introspection” (see Gennaro 1996: 19-21). The former is less sophisticated and only involves a brief conscious MET whereas the latter involves the use of reason and a more sustained inner-directed conscious thinking over time. I mainly have the former in mind throughout this chapter.

18. I should note here that Kriegel (2002: 525) does briefly mention a possible Brentanian account of introspection. However, numerous questions remain regarding how faithful such an account is to Brentano and how it would really help to save PSR (hence I leave the question mark in figure X.1).

Moreover, this and the previous objection to PSR are vividly illustrated in the writings of Hossack (2002, 2003). As an apparent supporter of PSR, Hossack frequently conflates first-order consciousness with introspection. He argues for an “identity thesis” defined as “each state of which one can be conscious is numerically identical with one’s introspective knowledge of the occurrence of that very state.” (2002: 163; emphasis added) The problem is that Hossack often seems to have in mind inner-directed conscious focus when speaking of “introspective awareness” and “self-knowledge” (cf. Hossack 2003: 196ff). Indeed, Hossack (2002: 174) goes so far as to hold that the “Identity Thesis says that every experience and every action is a conscious state, identical with knowledge of its own occurrence.” Thus, Hossack faces the following dilemma: Either such “self-knowledge” of a conscious state is outer-directed or it is inner directed. If the former, then he is not providing an analysis of M as part of an attempt to defend PSR; that is, M* would not be directed at M anyway. If the latter (as it appears), then he is not only conflating M* with introspective awareness, but clearly cannot justify identifying M* with conscious state M because M is an outer-directed state.

19. Moreover, Kriegel’s subsequent discussion of “three positions” (2003a: 116ff) does not exhaust the possibilities. As he knows (Kriegel, this volume), a fourth option is to treat M and M* as proper parts of a complex conscious state, even if we differ about whether or not M* is
itself conscious.

20. For example, as was mentioned in section 3.3, in order to explain the transition from first-order conscious states to introspective states. There are also numerous Kantian reasons to hold such a view ranging from the way that concepts are presupposed in experience to the role of episodic memory in consciousness (see Gennaro 1996, especially Chapters 3 and 9).

21. Brook and Raymont (forthcoming) also seem to have something related to the WIV in mind, especially in speaking of the global “representational base” of conscious states. I am also very sympathetic to their Kantian approach. However, it is unclear to me at this time just how similar or different our theories are.

22. To be fair, Kriegel has confirmed for me (in email correspondence) that he does not currently hold PSR as he did at the time of writing Kriegel 2003a. However, he does still seem to hold that M* is itself conscious, which was my main target in the previous section.

23. See, for example, Kriegel: this volume, note 34; 2004, note 24.

24. For more discussion on this point, see Gennaro 1996: section 6.6.

25. It seems to me that, for example, Van Gulick’s (2000, 2004) HOGS model is also arguably committed to an unconscious part of a global conscious state. The basic idea behind his “higher-order global states” theory is that “lower-order object states become conscious by being incorporated as components into the higher-order global states (HOGS)....The transformation from unconscious to conscious state is not a matter of merely directing a separate and distinct meta-state onto the lower-order state but of “recruiting” it into the globally integrated state...” (2004: 74-75) It is difficult to understand Van Gulick’s frequent use of the expression “implicit meta-intentionality” of HOGS unless we (at least, often) think of it as an unconscious part of a conscious whole. Much like my METs which are implicitly (i.e. unconsciously) intrinsic to the structure of complex conscious states, such meta-intentional parts of HOGS would presumably also not be conscious; hence, they are “implicit” and not “explicitly” contained in the structure of conscious experience. Van Gulick normally speaks somewhat neutrally about whether or not the meta-intentional parts of his HOGS are conscious or unconscious, such as when he frequently
says they are “built into” or “embedded in” the HOGS. This leads me to construe his HOGS as similarly containing an unconscious meta-psychological aspect. However, if Van Gulick means to suggest that such meta-intentionality is *consciously* part of one’s outer-directed phenomenal experience, then I disagree for the same reasons given in the previous section against PSR, and especially against the phenomenological argument (section 3.4). A few passages *might* suggest such an interpretation; for example, “Experience is always the experience *of a self* and *of a world of objects.*” (Van Gulick 2004: 81) This carries various Kantian considerations, to which I am very sympathetic, a bit too far in my opinion. It is one thing to say that M* is “built into” or “presupposed in” the structure of conscious experience, but quite another to hold that M* is itself a conscious part of a conscious state.


27. Kriegel (2004), Caston (2002: 777, note 54), Ken Williford (email communication, 2003), but recall also, on the other hand, that some accuse me of really holding some form of FO theory.

28. Thomas Metzinger (1995: 454) also alludes to the view that consciousness “is something that unifies or *synthesizes* experience.” He calls this process “higher-order binding” (HOB).


30. This point is echoed in Van Gulick’s (2000, 2004) discussion of Chris Hill’s “volume control hypothesis” whereby introspection is an active process in the sense that it often alters its lower-order mental object. I agree with Van Gulick and Hill here; however, they are clearly discussing *introspection* only. What is really needed, however, is *also* the view that *unconscious* METs (or HOTs) affect the nature of their target states in those cases of *first-order* conscious states.

31. Also, see section 6.3 in Gennaro 1996 where I discuss Daniel Dennett’s case of Chase and Sanborn in light of the WIV.

32. I lack the space here to address Droege’s more positive account, which she calls “second sense theory.”
33. Recall also (from section 2) that intrinsic theory need not rely solely on such phenomenological evidence in order to support the notion that consciousness is an intrinsic part of conscious states.

34. See Thomasson (2000: 205-206) for some discussion of this objection.

35. Many thanks to Uriah Kriegel, Robert Lurz, and Ken Williford for very helpful comments on an earlier draft of this chapter.

References:


Gennaro, R. (unpublished). The HOT theory of consciousness: between a rock and a hard place?


Kriegel, U. (this volume). The same-order monitoring theory of consciousness.


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


