CANADIAN JOURNAL OF PHILOSOPHY
Volume 32, Number 3, September 2002, pp. 293-330

Jean-Paul Sartre and the HOT Theory of Consciousness

ROCCO J. GENNARO
Indiana State University
Terre Haute, IN 47809
USA

Jean-Paul Sartre believed that consciousness entails self-consciousness, or, even more strongly, that consciousness *is* self-consciousness. As Kathleen Wider puts it in her terrific book *The Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind*, 'all consciousness is, by its very nature, self-consciousness.'[1] I share this view with Sartre and have elsewhere argued for it at length.[2] My overall aim in this paper is to examine Sartre's theory of consciousness against the background of the so-called 'higher-order thought theory of consciousness' (the HOT theory) which, in turn, will shed light on the structure of conscious mental states as well as on Sartre's theory of (self-) consciousness and reflection. Another goal of this paper is, following Wider, to show how Sartre's views can be understood from a contemporary analytic perspective. Sartre's theory of consciousness is often confusing to the so-called 'analytic Anglo-American' tradition, but I attempt to show how this

1 Kathleen Wider, *The Bodily Nature of Consciousness: Sartre and Contemporary Philosophy of Mind* (Ithaca: Cornell University Press 1997), 1. I will hereafter refer to this book as **BNC**. Wider also provides us with a very good sense of the tradition behind the view that consciousness is self-consciousness through an examination of Descartes, Locke, and Kant, in ch. 1.

2 Rocco J. Gennaro, *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness* (Amsterdam: John Benjamins Publishers 1996). This book will hereafter be abbreviated as **CSC**.

obstacle can be overcome against the backdrop of a specific contemporary theory of consciousness.

In Section I, I explain some key Sartrean terminology and in Section II, I introduce the HOT theory. Section III is where I argue for the close connection between Sartre's theory and a somewhat modified version of the HOT theory. That section of the paper is divided into four subsections in which I also address the relevance of Sartre's rejection of the Freudian unconscious and the threat of an infinite regress in his theory of consciousness. In Section IV, I critically examine what I call 'the unity problem,' which has mainly been raised by Kathleen Wider against Sartre. In light of Section III, I attempt to relieve some of Sartre's difficulties. In Section V, I critically examine a passage from *Being and Nothingness*[3] containing one of Sartre's main arguments for his belief that consciousness entails self-consciousness. In Section VI, I show how Sartre and the HOT theory can accommodate so-called 'I-thoughts' into the structure of conscious mental states with the help of Wider's view. Finally, in Section VII, I offer some concluding remarks.

I Sartre's Terminology and Basic Theory

Sartre divides reality into what he calls 'being-in-itself' and 'being-for-itself.' The in-itself (*en-soi*) refers to nonconscious parts of reality whereas the for-itself (*pour-soi*) refers to consciousness and, more specifically, to human self-consciousness. Being-in-itself 'is what it is' (BN 29) whereas the for-itself 'is not what it is and is what it is not' (BN 120, 127).[4] The 'being of consciousness does not coincide with itself in a full equivalence' (BN 120; cf. BN 153). The for-itself is directed outside itself and is that through which negation and nothingness enter the world. Following Husserl, Sartre urges that 'all consciousness ... is consciousness *of* something' (BN 11, 23). The key point here is the essentially intentional aspect

---

3 Jean-Paul Sartre, *Being and Nothingness*, Hazel E. Barnes, trans. (New York: Philosophical Library 1956). All future references to this work will be abbreviated **BN** in the text followed by the page number. The page references will be to the paperback edition. The passage I am referring to here is BN 11.

4 I will return briefly to Sartre's puzzling notion that consciousness violates the Law of Identity later, in Section IV. On this topic, however, also see BNC 43-53, 150-4. For a good discussion of the for-itself/in-itself distinction, see Joseph Catalano, *A Commentary on Jean-Paul Sartre's Being and Nothingness* (Chicago: University of Chicago Press 1974), 41-8.

of consciousness.[5] When I am in a conscious mental state, it is directed at or 'about' something else.[6]

Sartre distinguishes between *positional (or thetic) consciousness* and *non-positional (or non-thetic) consciousness*. Cumming tells us that 'an act of consciousness is "positional" or "thetic" when it asserts the existence of its object.'[7] Obviously related to the intentional nature of consciousness, the idea is that when one's conscious attention is focused on something else, one 'posits' the existence of an intentional object. On the other hand, one merely has 'non-positional' consciousness of 'anything that falls within one's field of awareness but to which one is not now paying attention' (BNC 41). Every act of consciousness, Sartre eventually argues, has both a positional and non-positional aspect in ways that will become clear later. Sartre also distinguishes between *pre-reflective (or unreflective or non-reflective) consciousness* and *reflective consciousness*. Suffice it to say for now that the former is basically outer-directed consciousness and the latter is inner-directed consciousness.[8] Pre-reflective consciousness is what Sartre and commentators (with Descartes in mind) refer to as the 'pre-reflective *cogito*' whereas Sartre initially defines reflection as 'a consciousness which posits a consciousness.'[9]

---

5 For much more on Sartre and intentionality, see Phyllis Sutton Morris, *Sartre's Concept of a Person: An Analytic Approach* (Amherst: University of Massachusetts Press 1975). For more background on Sartre and his predecessors (especially Husserl), see BNC 41-3; Catalano, *Commentary*, 4-13; and William Schroeder, *Sartre and his Predecessors: The Self and the Other* (London: Routledge and Kegan Paul 1984).

6 It is questionable, however, that *all* mental states are intentional in this sense. For example, pains are not 'about anything.' There are no 'pains that p' or 'pains about x.' On the other hand, we might of course agree that pains are still 'representational' in some sense of the term, e.g. directed at a part of my body.

7 Robert Denoon Cumming, *The Philosophy of Jean-Paul Sartre* (New York: Random House 1965), 51n.

8 My understanding of the secondary literature is that in her 'Key to Special Terminology' at the end of BN, Hazel Barnes should not have *equated* pre-reflective or unreflective consciousness with non-thetic or non-positional self-consciousness. Nor should she have *equated* reflective consciousness with thetic or positional self-consciousness.

9 Jean-Paul Sartre, *The Transcendence of the Ego*, Forrest Williams and Robert Kirkpatrick, trans. (New York: Hill and Wang 1957), 62. All future references to this book will be abbreviated in the text as **TE** followed by the page number. It should also be noted here that Sartre does eventually distinguish between pure and impure reflection, which I briefly address later in Section III.4. I am primarily concerned with pure reflection throughout this paper, but the basic definition of reflection from TE is sufficient for my immediate purposes.

With the above sketch of Sartrean concepts in place, let us introduce the HOT theory before examining Sartre's theory in much greater detail.

II The HOT Theory

In the absence of any plausible reductionist account of consciousness in nonmentalistic terms, the HOT theory says that the best explanation for what makes a mental state conscious is that it is accompanied by a thought (or awareness) that one is in that state.[10] The intuitive idea, shared by Sartre, is that when one is in a conscious mental state one is certainly aware that one is in it. The sense of 'conscious state' I have in mind is the same as Nagel's sense, i.e. there is 'something it is like to be in that state' from a subjective or first-person point of view.[11] When I am, for example, having a conscious visual experience, there is something it 'seems' or 'feels' like from my subjective perspective. But when a conscious mental state is a first-order world-directed state the higher-order thought (HOT) is not itself conscious; otherwise, circularity and an infinite regress would follow. Moreover, when the HOT is itself conscious, there is a yet higher-order (or third-order) thought directed at the second-order state. In this case, we have *introspection* which involves a conscious HOT directed at an inner state. When one introspects, one's attention is directed back into one's mind.

For example, what makes my desire to write a good paper a conscious *first-order* desire is that there is a (nonconscious) HOT directed at the desire. In such a case, my conscious focus is directed at the paper. When I introspect that desire, however, I then have a *conscious* HOT directed at the desire itself. Figure 1 summarizes the contrast between first-order conscious states and introspective states on the standard HOT theory.

I suggest that self-consciousness is simply having meta-psychological or higher-order thoughts, even when the HOT is not itself conscious. A higher-order thought is, of course, simply a thought directed at another mental state. I have therefore argued at length in CSC that consciousness entails self-consciousness, but it is important to note here that there are degrees or levels of self-consciousness, with introspection as its more complex form. Thus, all introspection involves self-consciousness, but

---

10 See David Rosenthal, 'Two Concepts of Consciousness,' *Philosophical Studies* **49** (1986) 329-59. I also defend the HOT theory at great length in CSC.

11 Thomas Nagel, 'What is it Like to be a Bat?' *Philosophical Review* **83** (1974) 435-50.

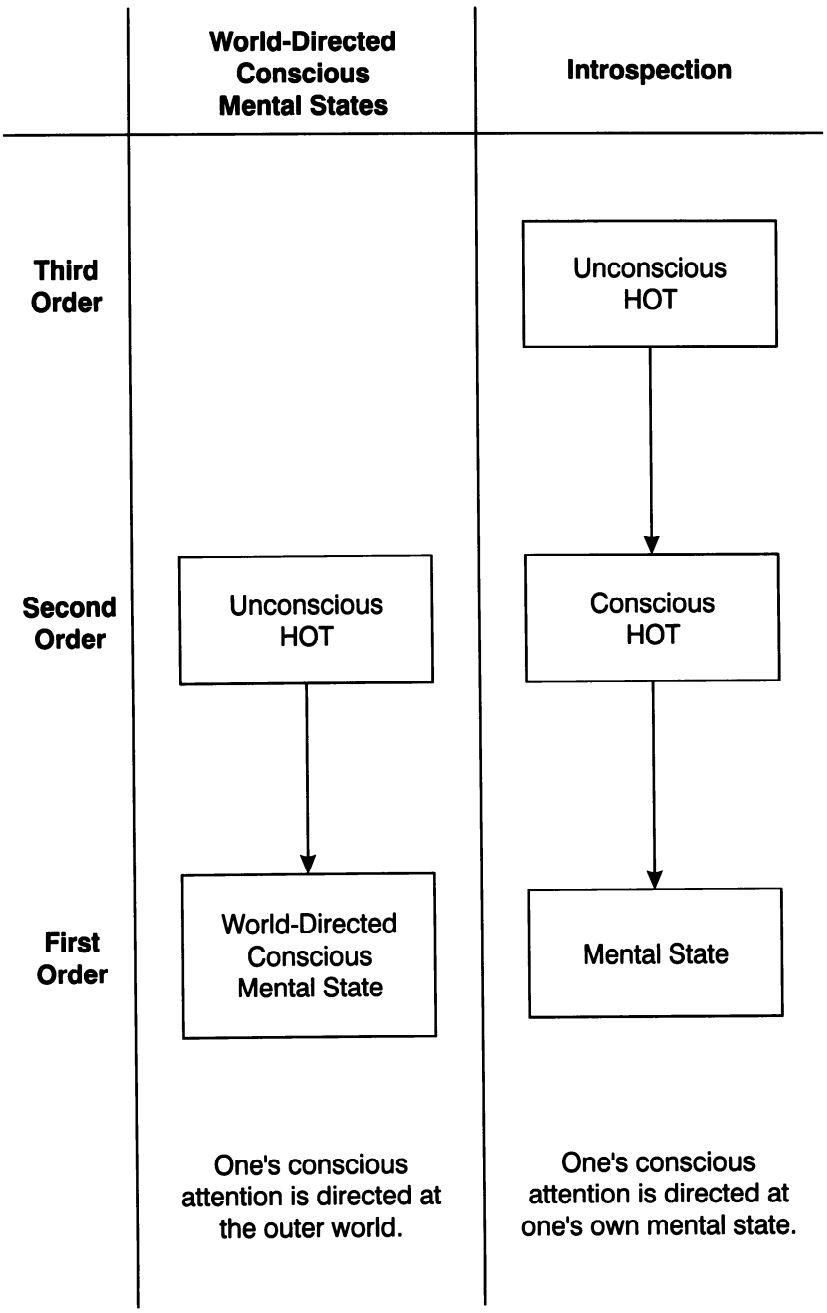

Figure 1. The HOT Theory of Consciousness

not necessarily vice versa. Some might still wonder why self-*consciousness* need not be *consciousness of* some *thing*. I offer two reasons here:[12]

(1) Few (if any) philosophers hold that self-consciousness is literally 'consciousness of a self,' especially since Hume observed that we are not aware of an unchanging or underlying self but only a succession of mental states. Thus the 'ordinary meaning' of 'self-consciousness' is somewhat open because the term does not wear its meaning on its sleeve. We are somewhat free to stipulate a meaning, though not of course in an entirely arbitrarily manner. It is clear from *The Transcendence of the Ego* that Sartre shares the view that there is no 'I' or 'self' standing behind one's sequence of mental states. There 'is no ego "in" or "behind" consciousness.'[13] This is Sartre's well-known rejection of Husserl's 'transcendental ego,' one of the two most important differences between Sartre and Husserl.[14] (2) Other philosophers have proposed even weaker definitions of 'self-consciousness.' For example, Van Gulick holds that it is simply the possession of meta-psychological information.[15] While I believe that that notion is too weak, my point here is only that my definition is not the weakest one in the literature. Owen Flanagan also recognizes a 'weaker' form of self-consciousness when he says that 'all subjective experience is self-conscious in the weak sense that there is something it is like for the subject to have that experience. This involves a sense that the experience is the subject's experience, that it ... occurs in her stream.'[16]

---

12 See CSC 17-18 for several additional reasons.

13 This is a quotation from the translator's introduction at TE 21. On this point see also Phyllis Berdt Kenevan, 'Self-Consciousness and the Ego in the Philosophy of Sartre,' in *The Philosophy of Jean-Paul Sartre*, Paul Schilpp, ed. (LaSalle: Open Court Press 1981), ch. 7.

14 The other major difference is Sartre's rejection of Husserl's 'bracketing' of the belief in the existence of outer phenomena.

15 Robert Van Gulick, 'A Functionalist Plea for Self-Consciousness,' *Philosophical Review* **97** (1988) 149-81. I argue that Van Gulick's notion of self-consciousness is too weak in CSC 147-51.

16 Owen Flanagan, *Consciousness Reconsidered* (Cambridge, MA: MIT Press 1992), 194

III Sartre and the HOT Theory

1. An initial problem: Sartre's rejection of the Freudian unconscious.[17]

The HOT theorist asks: what makes a mental state a conscious mental state? This is the fundamental question that should be answered by any theory of consciousness. The HOT theory says that what makes a mental state conscious is the presence of a suitable higher-order thought directed at it.[18] This allows, or even presupposes, that there can be unconscious mental states; that is, those mental states not accompanied by a HOT. However, Sartre explicitly rejects the existence of the Freudian unconscious, which would seem to rule out the existence of first-order nonconscious mental states. For example, Sartre says that 'pleasure cannot exist "before" consciousness of pleasure' (BN 14) and 'to believe is to know that one believes' (BN 114). Indeed, this is precisely what leads Sartre to address the problem of how so-called 'bad faith' (*la mauvaise foi*) is possible without presupposing an unconscious part of the mind. Sartre argues that postulating the Freudian unconscious would not even solve this paradox. Bad faith is basically 'lying to oneself' and is commonly treated as a form of self-deception.[19] In any case, Sartre was not

---

17 In this section I will use the terms 'unconscious' and 'nonconscious' interchangeably.

18 Of course, a full answer to the question 'What makes a higher-order thought "suitable"?' would require a lengthy digression that I cannot pursue here. One condition, for example, is that the HOT must be a momentary or occurrent state as opposed to a dispositional state. See CSC chapters 3 and 4 for my attempt at answering the above question. Moreover, the terminology can be a bit confusing. Sometimes the term 'thought' is used as a generic term covering all kinds of mental states, but it is also sometimes contrasted with 'perception.' For our purposes, we can think of the higher-order state as some kind of higher-order awareness. See CSC 95-101 for some discussion of this matter.

19 The topic of bad faith is a major issue in its own right that I cannot address here. For a small sample of the literature, however, see Robert Stone, 'Sartre on Bad Faith and Authenticity,' in *The Philosophy of Jean-Paul Sartre*, Paul Schilpp, ed. (LaSalle: Open Court Press 1981), ch. 10; Jeffrey Gordon, 'Bad Faith: A Dilemma,' *Philosophy* **60** (1985) 258-62; Joseph Catalano, 'Successfully Lying to Oneself: A Sartrean Perspective,' *Philosophy and Phenomenological Research* **50** (1990) 673-93; Ronald Santoni, *Bad Faith, Good Faith, and Authenticity in Sartre's Early Philosophy* (Philadelphia: Temple University Press 1995); Yiwei Zheng, 'Ontology and Ethics in Sartre's *Being and Nothingness*: On the Conditions of the Possibility of Bad Faith,' *The Southern Journal of Philosophy* **35** (1997) 265-87.

attempting to answer the above question in a way that would easily allow for unconscious first-order mental states.

Some commentators, however, have questioned Sartre's blanket rejection of the unconscious as apparently articulated in the section titled 'Bad Faith' in BN. Phyllis Sutton Morris notes how many Sartre scholars believe that Sartre eventually came to accept 'that there were "opaque" elements in our psychic lives.'[20] Even in BN, Morris points out that we find the following striking passage:

> the body is [the psyche's] substance and its perpetual condition of possibility.... It is this which is at the basis of the mechanistic and chemical metaphors which we use to classify and to explain the events of the psyche.... It is this, finally, which motivates and *to some degree justifies psychological theories like that of the unconscious*, problems like that of the preservation of memories. (BN 444, emphasis added)

We also find a surprising passage at BN 437 where Sartre seems to endorse a belief in unconscious pain when, for example, he writes, 'my reading "absorbs me" and when I "forget" my pain (which does not mean that it has disappeared since if I happen to gain knowledge of it in a later *reflective* act, it will be given as having always been there).' The parenthetical remark seems to suggest that the pain existed throughout his reading, even when he was not conscious of it.

Others have also argued that Sartre and Freud may not have been as far apart as is commonly believed. For example, Brown and Hausman question Sartre's commitment to the so-called 'translucency of consciousness' in comparing him to Freud.[21] Ivan Soll argues that even if Sartre has shown that Freud's postulation of an unconscious region of the mind does not resolve the paradox of bad faith, it does not follow that we must therefore reject the notion of unconscious mental processes. Soll explains, 'even if the postulation of the unconscious does not resolve the paradox of self-deception, it is quite clear that it was not postulated solely to resolve that paradox. Freud justified the postulation of the unconscious by claiming that it helped to explain several sorts of otherwise incomprehensible human behavioral phenomena, such as ...

---

20 Phyllis Sutton Morris, 'Sartre on the Self-Deceiver's Translucent Consciousness,' *Journal of the British Society for Phenomenology* **23** (1992) 103-19. The quotation is from page 115, but also see her footnote 28.

21 Lee Brown and Alan Hausman, 'Mechanism, Intentionality, and the Unconscious: A Comparison of Sartre and Freud,' in *The Philosophy of Jean-Paul Sartre*, Paul Schilpp, ed. (LaSalle: Open Court Press 1981), ch. 23.

dreams, memory, and various sorts of neurotic symptom formation.'[22] Part of Sartre's motivation to reject the unconscious no doubt also stems from his (very conscious!) desire to maintain his well-known and somewhat radical views on freedom and responsibility. As Soll puts it, 'Sartre also associates the postulation of a psychic unconscious with the introduction into the psychic realm of causal relationships and the freedom-threatening thesis of causal determinism so dear to Freud and so repugnant to him.'[23] It is certainly true that philosophers understandably tend to infer from a materialistic causal determinism to a lack of freedom and responsibility, at least in the robust sense that Sartre had in mind. Of course, Sartre could have instead argued that freedom enters the picture only at the level of *conscious* mentality while admitting the presence of unconscious mental states defined in terms of functional/behavioral roles. However, it is clear that, rightly or wrongly, Sartre believed that introducing an unconscious realm into his theory of consciousness would threaten a belief in freedom. After all, the belief in causally active unconscious mental states is frequently used by determinists in response to a wide variety of arguments for free will, such as the well-known 'argument from deliberation' whereby we infer that we really could have done otherwise from the first-person observation that we frequently deliberate over choices and then seem to be able to perform more than one action at a given time.

In any case, despite some very real questions regarding Sartre's views on unconscious mentality, it still seems unwise to hold that Sartre's position (especially in BN) can be made entirely consistent with this aspect of the HOT theory.[24] HOT theorists are united in their *unequivocal* acceptance of unconscious mental states. Nonetheless, Sartre was clearly still concerned to analyze and explain the structure of *conscious* mental states, and this is a desire he shared with HOT theorists. I will hereafter focus on this aspect of his theory. Even if one rejects the unconscious in some significant way, it still seems possible to offer an informative analysis of conscious mental states. This is where I disagree with Rosenthal when he argues that if we treat consciousness as an intrinsic property of

---

22 Ivan Soll, 'Sartre's Rejection of the Freudian Unconscious,' in *The Philosophy of Jean-Paul Sartre*, ch. 24, 586. Soll also argues that Sartre ignored or misunderstood some of Freud's more developed views.

23 Ivan Soll, 'Sartre's Rejection of the Freudian Unconscious,' 602

24 This is unlike, say, Leibniz who unambiguously believed in the unconscious and who also held a version of the HOT theory, I argue in 'Leibniz on Consciousness and Self-Consciousness,' in *New Essays on the Rationalists*, Rocco J. Gennaro and Charles Huenemann, eds. (New York: Oxford University Press 1999).

mental states, then conscious mental states will be simple and unanalyzable.[25] Descartes is the primary villain in his criticism of the view that consciousness is intrinsic and essential to mentality. Rosenthal defines an 'intrinsic' (as opposed to 'extrinsic') property as follows: P is an intrinsic property of x if x's having P does *not* consist in x bearing some relation R to *something else*.[26] But, as I have argued elsewhere,[27] even if consciousness is an intrinsic property of some or all mental states, it does not follow that the such mental states are simple or unanalyzable. In essence, Rosenthal has set 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' (the HOT), is an extrinsic property of mental states. But there is an informative third alternative that I call the 'wide intrinsicality view' (WIV) and that I will argue is very close to Sartre theory.

The WIV, in contrast to Rosenthal's version of the HOT theory, says that first-order conscious mental states are *complex* states containing both a world-directed mental state and a (nonconscious) meta-psychological thought (MET).[28] Conscious states are thus individuated 'widely.' As shown in figure 2, this alternative holds that consciousness is an intrinsic property of conscious states while also providing an *analysis* of conscious mentality. Moreover, *contra* Rosenthal's contention, such states are not simple, but rather are complex states with parts. On my view, the MET is a self-conscious state and so (like Sartre) even first-order conscious states are self-conscious. My conscious perception of the tree is accompanied by a (self-conscious) MET within the very same complex conscious state. Now when I *introspect* on 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

---

25 Rosenthal, 'Two Concepts of Consciousness,' 330, 340-8. See also David Rosenthal, 'A Theory of Consciousness,' Report No. 40 (1990) on MIND and BRAIN. Perspectives in Theoretical Psychology and the Philosophy of Mind (ZiF), University of Bielefeld, 22-4. A version of that paper is reprinted in *The Nature of Consciousness: Philosophical Debates*, Ned Block, Owen Flanagan, and Guven Guzeldere, eds. (Cambridge, MA: MIT Press 1997), ch. 46.

26 Rosenthal, 'A Theory of Consciousness,' 21-2.

27 CSC 21-4. I also show that Rosenthal mistakenly argues that intrinsicality entails essentiality and that extrinsicality entails contingency.

28 I will often use the expression 'meta-psychological thought' (MET) instead of 'higher-order thought' (HOT) because, on my view, the conscious rendering state is part of the first-order conscious state and so is technically not 'higher-order.'

noncomplex mental state which is the object of a higher-order conscious complex state (see figure 2). I believe that the WIV offers a neater and simpler alternative to the standard HOT theory.

There are a number of advantages to the WIV over Rosenthal's theory. I offer two here:[29]

(1) The WIV can very simply accommodate the intuitive belief that consciousness is an intrinsic property of mental states. From the first-person point of view, consciousness certainly seems to be an intrinsic feature of mental states (e.g. our visual perceptions). After all, consciousness does not seem to be an extrinsic property like 'being to the left of.' Even Rosenthal acknowledges that we should try to preserve this natural view if at all possible,[30] but then he rejects it for the reasons given above.

(2) The WIV also can explain the somewhat historically influential view that conscious mental states are, in some sense, directed at themselves. Conscious mental states are 'reflexive' or 'self-referential.' This is a view that Sartre held (as we will see more clearly later in this section) and it is also articulated by Brentano, who believed that conscious mental states are secondarily directed at themselves.[31] Brentano did not think that we could distinguish, say, the mental act of perceiving some object from the mental act of thinking that one is perceiving that object. Of course, strictly speaking, a conscious mental state is not self-referential in the sense that it is directed back at itself. There is instead an inner reflexivity *within* the complex conscious state such that the MET is directed at a part of the state of which it is part. In a similar way, Marjorie Grene says on Sartre's behalf that 'consciousness, to be consciousness, must be self-directed and self-contained.'[32]

I suppose it is open to Rosenthal to respond that what is reflexive is the amalgam of the HOT and the target (i.e. lower-order) state. Perhaps he could simply admit that, even on his version of the HOT theory, we sometimes refer to that amalgam as the 'conscious state.' If this is meant as a shift of position to the WIV, then such a reply would be welcome. The problem for Rosenthal, however, is that such an admission seems inconsistent with much of what he says against the idea that consciousness is an intrinsic property of conscious mental states. As we saw above, his considered position is that consciousness is extrinsic to the target

---

29 I describe five advantages in CSC 26-30.

30 Rosenthal, 'Two Concepts of Consciousness,' 331

31 Franz Brentano, *Psychology from an Empirical Standpoint* (New York: Humanities Press 1973 [1874])

32 Marjorie Grene, *Sartre* (New York: New Viewpoints 1973), 121

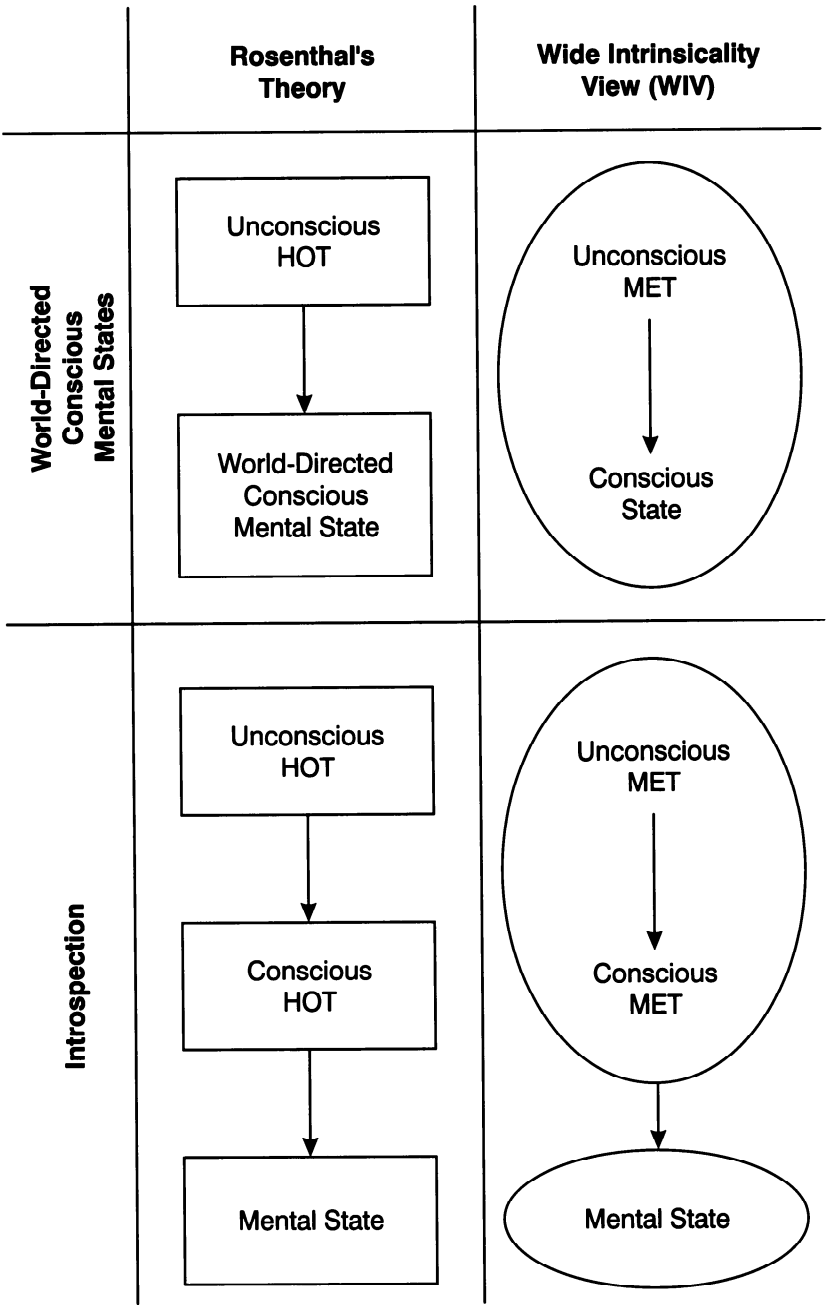

Figure 2.

state and he is frequently at great pains to demonstrate that Brentano's view is mistaken.[33]

In any case, we can see that even if Sartre did not believe in nonconscious mental states, it is crucial to separate that view from his analysis of consciousness. There are two different questions here: (1) Are mental states essentially conscious? and (2) Are conscious mental states essentially self-conscious? It certainly seems possible to hold that the answer to (1) is yes while the answer to (2) is also yes. Most of us would answer no to (1), but I would also answer yes to (2). It is clear, however, that these questions are independent and an answer to one need not logically lead us to any particular answer to the other. So mental states might not be essentially conscious, but conscious mental states can still be essentially self-conscious. In much of his discussion on bad faith, Sartre himself seems to have been unnecessarily concerned that answering no to (1) would cause problems for an affirmative answer to (2). This is presumably what Wider means when, in examining possible counterexamples to the thesis that consciousness entails self-consciousness, she says she will 'bypass the central counterexample Sartre focuses on in [BN] — the Freudian unconscious' (BNC 93). It is not clear to me why either Sartre or Wider would view the rejection of the unconscious as a counterexample to the thesis that conscious states are self-conscious.

Let us now proceed to another aspect of the HOT theory addressed by Sartre.

2. *The infinite regress.*

As I mentioned in Section II, the HOT theorist avoids definitional circularity and an infinite regress by explaining that the HOT is not itself

---

33 See also, for example, much of Rosenthal's argument in 'Thinking That one Thinks,' in *Consciousness*, Martin Davies and Glyn W. Humphreys, eds. (Oxford: Blackwell 1993), 197-223. Among other things, Rosenthal argues there that HOTs cannot be intrinsic to the conscious state because it would seem almost contradictory to have, for example, a doubt that it is raining but also an affirmative thought that I am in such a state. In other words, the very same conscious state cannot have parts with more than one mental attitude, e.g. doubting and affirming (assertoric). However, it is unclear why this should be so and that there is really a problem here for the WIV. In such a case, we would have a first-order conscious doubt directed at the weather accompanied by a MET of the form 'I (nonconsciously but assertorically) think that I am doubting it is raining.' The MET affirms the doubt and that affirmation is what makes the lower-order doubt conscious. Thus the complex conscious state is still a first-order world-directed conscious doubt, albeit with an assertoric meta-psychological component.

conscious when one has a first-order conscious state. Otherwise, we would be explaining consciousness by appealing to consciousness, which is circular. Moreover, we would have an infinite regress because for every conscious state there would have to be a higher-order conscious state and so on *ad infinitum*.[34] So, for example, on the WIV, the MET is a nonconscious part of a first-order conscious mental state. Sartre interestingly noticed a similar problem, but instead of straightforwardly responding in like manner, he first says the following in TE:

> All reflecting consciousness is, indeed, in itself unreflected, and a new act of the third degree is necessary in order to posit it. Moreover, there is no infinite regress here, since a consciousness has no need at all of a reflecting consciousness in order to be conscious of itself. It simply does not posit itself as an object. (TE 45)

And then in BN Sartre puts it as follows:

> Either we stop at any one term of the series — the known, the knower known, the knower known by the knower, *etc.* In this case the totality of the phenomenon falls into the unknown; that is, we always bump up against a non-conscious reflection and a final term. Or else we affirm the necessity of an infinite regress (*idea ideae ideae, etc.*), which is absurd.... Are we obliged after all to introduce the law of this [knower-known] dyad into consciousness? Consciousness of self is not dual. If we wish to avoid an infinite regress, there must be an immediate, non-cognitive relation of the self to itself. (BN 12)

What are we to make of these passages? A full answer to this question will not be entirely clear until the end of Section IV. However, to anticipate some of that discussion, we can see that in the TE quote Sartre is first recognizing that when there is 'reflecting' (i.e. introspective) consciousness, there must be 'a new act of the third degree.' This is reminiscent of the HOT theorist's contention that a third-order state is necessary for introspection. But there is no infinite regress because 'a consciousness has no need of a reflecting consciousness in order to be conscious of itself' which can be taken as meaning 'conscious mental states need not have a reflective (or introspective) state directed at it in order to be self-conscious.' The idea that a conscious mental state need not be accompanied by introspection is certainly the view of any HOT theorist. One can, for example, have an outer directed conscious mental state of a table without being *introspectively* conscious of one's own

---

34 For an example of this type of error, see Peter Carruthers, 'Brute Experience,' *Journal of Philosophy* **86** (1989) 258-69. See my reply to Carruthers in 'Brute Experience and the Higher-Order Thought Theory of Consciousness,' *Philosophical Papers* **22** (1993) 51-69.

perception. After all, one's *conscious* attention cannot be directed both at the table and at one's own mental state.

However, Sartre still must account for the non-positional awareness in pre-reflective consciousness that I described briefly in Section I. For example, Sartre holds that 'every positional consciousness of an object is at the same time a non-positional consciousness of itself' (BN 13). In the TE passage, Sartre is making the point that such non-positional consciousness of itself 'does not posit itself as an object.' As I will argue in the next subsection, this is tantamount to holding the WIV where the non-positional self-consciousness is part of the conscious mental state. Such self-consciousness is *not separate from* the mental state it is directed at, and this is why Sartre says that it 'does not posit an object.' So despite Sartre's claim that all consciousness is consciousness *of* something, he is apparently saying that, when it comes to such self-consciousness, it does not really posit an object, or at least not a *distinct* object. After all, he does call it '*non-positional* self-consciousness' (BN 26, emphasis added). This is also why he uses the 'of' [*de*] in parentheses merely out of 'grammatical necessity' when speaking of such non-positional self-consciousness (of) self.[35]

Similarly, in the BN quotation, Sartre avoids the regress by first rejecting any separation between such self-consciousness and the world-directed conscious state; that is, by rejecting the 'knower-known' dyad. Sartre then makes it clear that a conscious state, and so self-consciousness, is 'not dual.' And so, again, 'if we wish to avoid an infinite regress, there must be an immediate, non-cognitive relation of the self to itself' (BN 12). I will return to this last statement in Section IV, but we can already see how Sartre is trying to make room for some kind of meta-psychological awareness *which is part of* each conscious mental state.

But why not just call such meta-psychological awareness a 'nonconscious thought (or awareness)'? Sartre's reluctance to call the 'non-positional awareness' in pre-reflective consciousness a 'nonconscious thought' is perhaps partly due to his rejection of the first-order unconscious. However, I suggest that we must interpret Sartre as logically committed to the existence of nonconscious METs. What else is a 'non-positional self-awareness' except some kind of nonconscious meta-psychological mental state? I cannot understand it any other way. Such an awareness is clearly a *mental* state *of some kind*. Indeed, this seems to be the standard interpretation offered by numerous commentators. For example, Thomas Busch tells us that 'unreflective consciousness intends or posits an object other than itself and is simultaneously non-position-

35 See BN 14. Also see Catalano, *Commentary*, 32-3; and BNC 86-7 and note 14.

ally self-*aware*,'[36] and Peter Caws explains that Sartre insists on 'the necessary accompaniment of every act of consciousness with a state of unreflective self-*awareness*.'[37] Furthermore, *either* 'non-positional self-awareness' is conscious (which is absurd and would lead to the infinite regress) *or* it is a nonconscious mental state of some kind (which is therefore the only viable alternative). One's *conscious* mind cannot be directed *both* at outer objects *and* at one's own mind at the same time, and so the non-positional self-awareness that accompanies one's pre-reflective (positional) consciousness of outer objects must be nonconscious.

3. Sartre and the WIV: Additional textual and commentator support.

To support further the close connection between Sartre's theory and the WIV, let us look extensively at some primary texts and secondary sources. My aim is to show that Sartre held a view very much like the WIV, though he used his own unique terminology. Figure 3 is a side by side comparison of the two theories.

To remind us of Sartre's overall position, Hazel Barnes says the following:

> by nature all consciousness is self-consciousness.... When I am aware of a chair, I am non-reflectively conscious of my awareness. But when I deliberately think of my awareness, this is a totally new consciousness; and here only am I explicitly positing my awareness or myself as an object of reflection. The pre-reflective cogito is a non-positional self-consciousness. (BN xi)

As we have already seen, this is in much the same spirit as the HOT theory. Let us now first look at the structure of *pre-reflective* conscious mental states. Recall Sartre's claim that 'every positional consciousness of an object is at the same time a non-positional consciousness of itself' (BN 13). Catalano, for example, also explains, 'consciousness is *directly* an awareness of something other than itself and simultaneously and indirectly an awareness of itself, as when we are absorbed in a book, we are directly aware of reading and indirectly aware of *ourselves* as read-

---

36 Thomas W. Busch, 'Sartre's Use of the Reduction: *Being and Nothingness* Reconsidered,' in *Jean-Paul Sartre: Contemporary Approaches to His Philosophy*, Hugh Silverman and Frederick Elliston, eds. (Pittsburgh: Duquesne University Press 1980), 19, emphasis added

37 Peter Caws, *Sartre* (London: Routledge and Kegan Paul 1979), 55

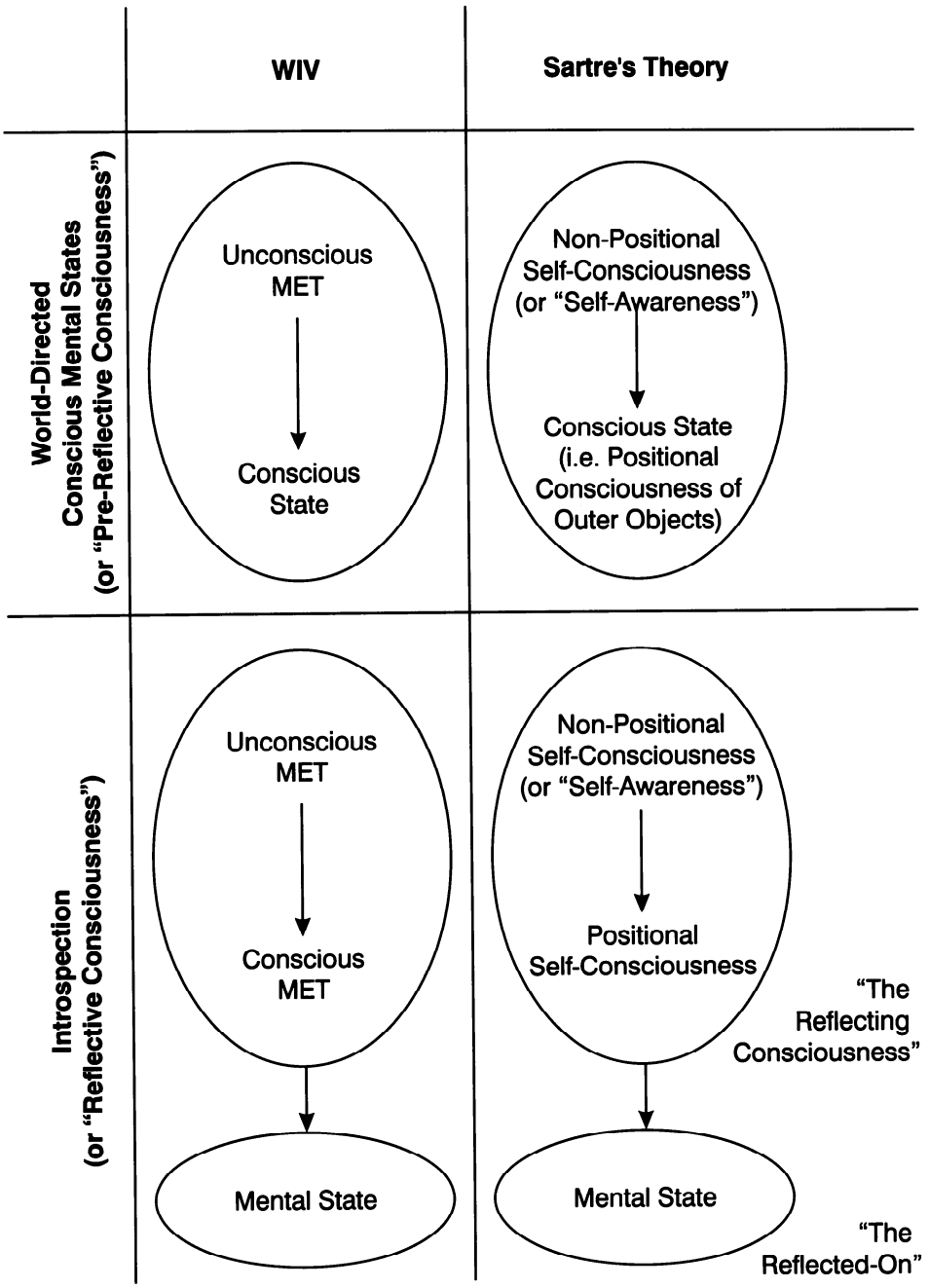

Figure 3.

ing.'[38] And Morris says that 'we are usually directly aware of an [outer] object, but not explicitly of ourselves being aware of (desiring, thinking of, etc.) of that object.'[39] Thus when our consciousness is outer-directed we are not explicitly (i.e. reflectively) aware of the lower-order state, but we are implicitly (i.e. non-positionally) aware of ourselves being aware of the outer object. Morris also explains that 'our most common conscious activities are prereflective or unreflective — not in the sense of being thoughtless, but rather in the sense of being explicitly directed toward objects other than ourselves and our actions.... It is only when someone takes his own conscious activities as the object of his attention that he has begun to reflect in Sartre's sense.'[40]

To bring this even closer to the WIV version of the HOT theory, consider the following passage:

> The immediate [pre-reflective] consciousness which I have of perceiving does not permit me either to judge [it].... It does not *know* my perception, does not *posit* it; all that there is of intention in my actual consciousness is directed toward the outside, toward the world. In turn, this spontaneous consciousness of my perception is *constitutive* of my perceptive consciousness. (BN 12-13)

Similarly, Sartre says 'Consciousness (of) pleasure is *constitutive* of the pleasure as the very mode of its own existence' (BN 14, emphasis added). So Sartre is recognizing, as he did in avoiding the infinite regress, that the non-positional self-consciousness is part of the lower-order state. It is 'constitutive' of the first-order conscious state.

Perhaps even more striking is the following passage:

> it is the very nature of consciousness to exist "in a circle".... We understand now why the first consciousness of consciousness is not positional; it is because it is one with the consciousness of which it is consciousness. (BN 13-4)

Looking at figure 3 one can see how this accords with pre-reflective consciousness. The non-positional (self-) consciousness is the 'first consciousness of consciousness' and 'it is one with the consciousness of which it is conscious [i.e. the world-directed conscious state].' I believe that this is what Sartre meant by saying that consciousness exists 'in a circle.'

It is also clear that Sartre views the non-positional self-awareness within pre-reflective consciousness as a form of self-consciousness. Oth-

---

38 Catalano, *Commentary*, 33

39 Morris, *Sartre's Concept of a Person*, 31

40 Morris, 'Sartre on the Self-Deceiver's Translucent Consciousness,' 108

erwise, he would not be able to claim that all consciousness is self-consciousness. In the same way, on my view, the nonconscious MET in a world-directed conscious state is a form of self-consciousness. Recall that in Section II, I defined self-consciousness as simply having meta-psychological thoughts, even when the MET is not itself conscious. Sartre makes numerous references to self-consciousness in much the same spirit: 'This [non-positional] self-consciousness we ought to consider not as a new consciousness.... [it is] a *quality* of the positional consciousness' (BN 14). He speaks of 'non-positional self-consciousness' (BN 26), 'non-thetic self-consciousness' (BN 120), and says that 'pre-reflective consciousness is self-consciousness' (BN 123). Thus Catalano tells us that 'consciousness, by its very being as an awareness, is *pre-reflectively* a (self-) consciousness.'[41]

Moving up the line, so to speak, to *reflective* consciousness, it is widely acknowledged that part of Sartre's motivation for distinguishing reflective and pre-reflective consciousness is to separate himself from Descartes who Sartre believed ignored the pre-reflective *cogito* in his methodological doubt. The pre-reflective cogito is actually a necessary condition of reflective consciousness. In the same way, on the WIV, one cannot have reflective (i.e. introspective) states without first having a first-order state. It is precisely when the first MET *becomes conscious* that one is reflecting. Thus, we find Sartre saying, 'it is the non-reflective consciousness which renders the reflection possible; there is a pre-reflective cogito which is the condition of the Cartesian cogito' (BN 13). Moreover, 'the unreflected has the ontological priority over the reflected because the unreflected does not need to be reflected in order to exist, and because reflection presupposes the intervention of a second-degree consciousness' (TE 57-8). In short, reflective states presuppose pre-reflective states, but not vice versa. As Busch puts it, 'the unreflected act, must be differentiated from the subsequent, or secondary, operation whereby a reflecting consciousness comes to bear in an objective way upon the unreflective or pre-reflective consciousness.'[42]

Recall that in Section I we noted how Sartre defined reflection as 'a consciousness which posits a consciousness'(TE 62). He also says that reflection is 'an operation of the second degree ... performed by a[n act of] consciousness *directed upon consciousness*, a consciousness which takes consciousness as an object' (TE 44). Thus the higher-order (i.e.

---

41 Catalano, *Commentary*, 32

42 Thomas Busch, *The Power of Consciousness and the Force of Circumstances in Sartre's Philosophy* (Bloomington: Indiana University Press 1990), 5

second-order) reflecting consciousness 'posits' a lower-order consciousness, as is shown in figure 3. We can call the higher-order complex state 'the reflecting consciousness' and the lower-order state 'the reflected-on.' And so reflection, for both Sartre and the WIV, involves a reflecting consciousness directed at an inner reflected-on object (i.e. mental state). So 'the reflecting consciousness posits the consciousness reflected-on, as its object. In the act of reflecting I pass judgment on the consciousness reflected-on' (BN 12-13). And also 'reflection is the for-itself conscious *of* itself. As the for-itself is already a non-thetic [i.e. non-positional] self-consciousness, we are accustomed to represent reflection as a new consciousness ... directed on the consciousness reflected-on' (BN 212).

Now one might wonder in looking at figure 3: does Sartre really also hold that the higher-order ('reflecting') state is a complex state in the same way as the WIV holds? I think the answer is yes, partly because it is the only way for Sartre to be consistent in his analysis. Much like the WIV, what goes for lower-order conscious states must also go for higher-order conscious states. Morris agrees when she says that 'the act of reflection exhibits the same internal dimensions of focused/non-focused awareness as do prereflective acts of consciousness.'[43] These 'internal dimensions' are present in the complex higher-order reflecting state in the same way as they are in a pre-reflective conscious state. Thomas Busch, in a similar way, explains:

> The reflective consciousness intends in a positional manner the pre-reflective consciousness. Pre-reflective consciousness intends in a positional manner some object other than itself, *but both levels, in addition ... are non-positionally self-aware.... It is important to note that the implicit, or non-positional, self-awareness exists on both levels, pre-reflective and reflective.*[44]

> Reflective consciousness intends or posits an object, but in its case the object is another act of consciousness. *The reflective consciousness is also non-positionally self-aware.... The basic structure of consciousness, in either case, is the same: awareness of an object and simultaneous awareness of being aware of the object.*[45]

Such quotations fit nicely with my comparison of Sartre's theory of reflective consciousness and the explanation offered by the WIV. The

---

43 Morris, 'Sartre on the Self-Deceiver's Translucent Consciousness,' 108

44 Thomas Busch, *The Power of Consciousness and the Force of Circumstances in Sartre's Philosophy*, 7, emphasis added

45 Thomas Busch, 'Sartre's Use of the Reduction: *Being and Nothingness* Reconsidered,' 19, emphasis added

higher-order reflecting state has the same internal structure as a first-order conscious state, but in the former case its object is another mental state whereas in the latter case it is an outer object.

Thus, to summarize, it is clear that Sartre views 'non-positional self-awareness' as a form of self-consciousness since he believed that first-order outer-directed conscious states are also self-conscious states. Like a HOT theorist, however, he also recognized a higher-order form of self-consciousness which he called 'reflection' in BN (instead of 'introspection'). So when one is in a first-order conscious mental state, one has a complex state such that one is positionally aware of an outer object but also non-positionally aware of that awareness. When one is in a second-order reflective state, one has a complex higher-order conscious 'reflecting' state directed at (or positionally aware of) a first-order 'reflected-on' state of consciousness. In this case, the complex second-order state is also constituted by both positional awareness (of the lower-order state) and a non-positional awareness of itself.

4. *Two types of reflection.*

In order to complete the discussion up to this point it is necessary to mention briefly Sartre's distinction between *pure* and *impure* reflection. As I noted at the end of section one, my use of the term 'reflection' until now has been short for what Sartre calls 'pure reflection.' This is because it is pure reflection that is closer to the notion of 'reflecting on one's current states of mind.' Pure reflection is more of an 'immediate reflection' on ourselves or on the present that has just been made past.[46] On the other hand, impure reflection is more on our 'remote past' and is a 'more deliberate and, therefore, cognitive reflection.' Impure reflection is the 'reflection on ourselves as a succession of states.'[47] Although the structure of both pure and impure reflection is the same in the sense that they both involve conscious higher-order states directed at lower-order states, impure reflection can also take already past mental states as objects and so can be described as a 'knowledge of myself' that 'fixes the for-itself as an in-itself.'[48] On the other hand, 'in pure reflection consciousness attempts to be present to itself as a present moment of consciousness' (BNC 79). In Sartre's own words, 'pure reflection [is] the

---

46 See, e.g., Catalano, *Commentary*, 126, 129-130.

47 The previous three brief quotations are from Catalano's *Commentary* 130, 126, 130 respectively.

48 Catalano, *Commentary*, 130

simple presence of the reflective for-itself to the for-itself reflected-on' (BN 218). On the other hand, impure reflection 'constitutes the succession of psychic facts or *psyche*' (BN 223), though this 'includes pure reflection as its original structure' (BN 224).

I have also distinguished between two forms of introspection (i.e. reflection).[49] *Momentary focused introspection* only involves a brief conscious HOT while *deliberate introspection* involves the use of reason and a more sustained inner-directed conscious thinking over time. Sometimes we consciously think to ourselves in a *deliberate* manner, e.g., in thinking about our philosophical views or our life goals. But one might also consciously think about a mental state without deliberating in any way, e.g., *momentarily* think about a memory or briefly consciously focus on a pain or emotion. In these cases, one is not engaged in deliberation or reasoning. Many animals, for example, seem capable of this kind of introspection even if they cannot deliberate, though both types of introspection still involve having conscious HOTs directed at lower order mental states.

It is tempting to *identify* Sartre's pure reflection with my momentary focused introspection and his impure reflection with my deliberate introspection, but that would be a mistake mainly because it seems to me that we can and often do deliberate about our *current* states of mind. We might, for example, be engaged in reflective deliberate examination of our own current philosophical beliefs. While Sartre is right that we may reflect on our 'past' succession of mental states, he seems to reserve such deliberation to impure reflection. However, deliberate introspection can be both pure and impure.

IV The Unity Problem

In this Section, I apply the results from Section III to what I will call 'the unity and separation problem,' or 'the unity problem' for short, which is vividly presented by Kathleen Wider.[50] Although I am somewhat in agreement with Wider's criticism of Sartre, I show how we can defend Sartre to at least some degree.

So what is the unity problem? In short, it is this: how can a lower-order state be *the object of* a meta-psychological state while being the very same

---

49 See CSC 19-21.

50 This is related to what Wider calls her 'internal critique' in BNC ch. 3, but the unity problem is most forcefully argued in Wider's 'Through the Looking Glass: Sartre on Knowledge and the Pre-reflective *Cogito*,' *Man and World* **22** (1989) 329-43. I will hereafter refer to this paper as **TLG**.

state (particularly since Sartre also maintains that there is *knowledge* of the lower-order state)? I believe we have seen how the WIV can help us understand how this is possible *on the pre-reflective level* if we view the MET as *part of* the complex lower-order conscious state. We can at least see how one mental state can be 'directed at' another while still being part of the same (complex) state, even though Sartre himself clearly struggles with the details of his theory on this point. While this problem is perhaps even more apparent on the reflective level, Wider argues that it is present on the pre-reflective level as well. She points out, for example, that Sartre says 'to believe is to *know* that one believes' (BN 114). The problem is that, for Sartre, consciousness is not the same as knowledge. One can be conscious of x and not have knowledge of x according to Sartre's use of these terms. 'Knowledge is nothing other than the presence of being to the For-itself.' (BN 295) Knowledge involves a *separation* of the knower and the known object. Thus, even on the pre-reflective level, Sartre cannot maintain both that non-positional self-consciousness is knowledge of the 'lower part' of the first-order conscious state and then maintain that there is no separation between the two: ' ... presence to always implies duality, at least a virtual separation. The presence of being to itself implies a detachment on the part of being in relation to itself ... But if we ask ourselves ... *what it is* which separates the subject from himself, we are forced to admit that it is *nothing*.' (BN 124) As Wider puts it, ' ... to know an object requires that one not be the object. Knowledge involves negation, a separation of the knower from the known ... Presence involves duality and separation ... ' (TLG 338) But the WIV does allow for a 'duality' within the pre-reflective level, just not for any *literal* 'separation' between the MET [i.e. non-positional self-consciousness] and the other part of the complex conscious state. Perhaps this is what Sartre meant by a 'virtual separation;' namely, a separation *within* a complex conscious state.

Although the WIV can help Sartre to some degree, Wider is correct that he continues to have a problem. Even if I am right in interpreting Sartre as holding the WIV on the pre-reflective level, the unity problem still remains with respect to the above remarks about knowledge. I believe that Sartre did have something like the WIV in mind, but was simply struggling with how to characterize the relationship between the parts of the complex pre-reflective conscious state. Speaking of a 'duality within a unity' makes sense on the WIV. There is also sense to be made of the claim that there is 'nothing' between the parts, since they are part of the same conscious mental state. So some of the tension in Sartre's thought can be relieved by my analysis. However, Sartre still runs into serious problems and ambiguities when describing whether or not non-positional self-consciousness is a form of *knowledge*. It is perhaps understandable, however, *why* Sartre struggled so much with this problem.

Like the WIV, he is trying to make sense of a 'directedness' or 'aboutness' *within* the pre-reflective level, but, in doing so, he sometimes characterizes that relation as a kind of knowledge.[51]

Recall that we confronted a version of this problem in section III.2 while addressing the threat of an infinite regress. Sartre explained that 'if we wish to avoid an infinite regress, there must be an immediate, non-cognitive relation of the self to itself' (BN 12). Indeed, Wider rightly emphasizes the term 'non-*cognitive*' in this quotation to support her contention that, at least in the Introduction to BN, Sartre did not want there to be a knowledge relation between the parts of pre-reflective consciousness (i.e., between non-positional self-consciousness and its 'object'). Although the French term Sartre uses for 'cognitive' [*cogitif*] is not identical with his usual term for 'knowledge' [*connaissance*], it does seem reasonable to take '*non cogitif*' as at least *implying* 'not knowledge.' In comparison to the HOT theory, it is worth briefly digressing here to mention the importance of the term 'immediate' in the BN 12 quotation. On the HOT theory, the relation between the HOT (or MET) and the lower-order conscious state must be immediate. The meta-awareness of the lower-order state must be *direct*; that is, the person is not aware of the conscious state in virtue of being aware of any other state. This is much like Sartre's definition of 'immediacy,' which, he says, 'is the absence of any mediator; that is obvious, for otherwise the mediator alone would be known and not what is mediated' (BN 247). The main reason for this condition on the HOT theory is to avoid alleged counter-examples purporting to show that one can have a HOT directed at one's own mental state of, say, anger at one's boss, but still not *feel* the anger. This could happen if, for example, one is confronted by another's observation of one's behavior or one is told so by a trusted psychotherapist. But such HOTs are not directly or immediately aware of the mental state of anger; rather, they are *inferred* by virtue of being immediately aware of my behavior or someone else. This is also why a HOT theorist should not allow the conscious rendering HOT to arise *via inference* or as a result of indirect evidence. It is wisest for a HOT theorist to hold that the HOT (or MET) must meet this so-called 'noninferentiality condition' in order for the lower-order state to be conscious.[52]

---

51 Wider does point out that Sartre recognizes the problem to some extent, but he is much too quick to dismiss it. See TLG 334.

52 For more on this condition, see CSC 84-7, and Rosenthal, 'Two Concepts of Consciousness,' 335-6.

Let us return to the unity problem. I believe that it arises in even more dramatic fashion at the reflective level despite some help from the WIV and our figure 3 comparison. Sartre seems to contradict himself on the question of whether or not there is a knowledge relation between the higher-order (complex) reflecting state and the lower-order reflected-on state. For example, he says that 'the reflected-on is not wholly an object but a *quasi-object* for [pure] reflection' (BN 218, emphasis added). What does a 'quasi-object' mean in this context? And to the question 'Is [pure] reflection a form of knowledge?' Sartre first tells us that 'reflection is a knowledge; of that there is no doubt. It is provided with a positional character; it affirms the consciousness reflected-on' (BN 218). But he then soon after contradicts or at least weakens that claim by saying that the reflective 'does not detach itself completely from the reflected-on' and 'reflection is a *recognition* rather than knowledge' (BN 218, emphasis added). It is never made clear just how such 'recognition' differs from 'knowledge.' Thus on both levels, Wider summarizes the tension by explaining that

> in order to defend his claim that there is duality even in the unity of pre-reflective consciousness, [Sartre] illegitimately introduces cognitive elements into the discussion of consciousness at that level [and] in order to reassert the unity of consciousness at the level of pure reflection, he weakens and at times abandons his claim that pure reflection is knowledge. (TLG 340)

Indeed, Wider argues that the problem is even more serious when we consider that it also infects Sartre's characterization of *impure* reflection. Sartre *unambiguously* wants to maintain that impure reflection in a form of knowledge since this occurs when 'the reflective consciousness tries to take a point of view on the consciousness reflected on and thus attempts to view the reflected consciousness as an object' (BNC 77). Sartre describes impure reflection as the attempt 'to apprehend the reflected-on *as in-itself* in order to make itself be that in-itself which is apprehended' (BN 224, emphasis added). But Sartre then has difficulty distinguishing pure from impure reflection when he characterizes pure reflection in terms of knowledge. And then when he treats pre-reflective self-consciousness as a form of knowledge, Sartre has difficulty separating it from either type of reflection. So according to Wider:

> [Sartre's] account fails on the level of pure reflection because [he] offers no account of self-consciousness on this level that succeeds in distinguishing pure reflection from the self-consciousness of pre-reflective consciousness without at the same time causing the collapse of his distinction between pure and impure reflection. (BNC 91)

But even if Sartre's view is hopeless, or even contradictory, on the knowledge aspect of all this, I suggest that our analysis of the structure

of reflective states (in terms of figure 3) can shed some light on the nature of this problem and relieve some of the tension. Given his definition of knowledge, Sartre probably should have said that knowledge appears *neither* at the pre-reflective *nor* at the pure reflective level, and then left knowledge only for impure reflection. Sartre had something like the WIV in mind, but was not very clear about how to characterize the relationship between the complex higher-order reflecting state and the lower-order reflected-on state, particularly since knowledge involves a 'separation.' In an attempt to stress the unity of reflective states, Sartre even flirts with characterizing reflective consciousness as something like what I have in figure 4; that is, treat reflection as one very complex conscious state ('one big circle') with parts including both the reflecting and reflected-on consciousnesses. So, like the pre-reflective level, there is a kind of 'separation' or 'duality,' but still within a 'unity.' This would also again help to explain his ambiguous treatment of the knowledge relation.

Figure 4 may indeed represent his considered mature view, but we can also find textual support for it as early as in TE:

> there is an *indissoluble unity of the reflecting consciousness and the reflected consciousness* (to the point that the reflecting consciousness could not exist without the reflected consciousness). But the fact remains that we are in the presence of a synthesis of two consciousnesses, one of which is consciousness *of* the other. (TE 44, emphasis added)

> *reflection and reflected are only one ... and the interiority of the one fuses with that of the other*. To posit interiority before oneself, however, is necessarily to give it the load of an object. This transpires as if interiority closed upon itself and proffered us only its outside; as if one had to "circle about" it in order to understand it. (TE 84, emphasis added)

In BN Sartre also speaks of the reflective consciousness and 'its absolute unity with the consciousness reflected-on' (BN 212) and says that 'reflection is *one being*, just like the unreflective for-itself, not an addition of being' (BN 215). So, for example, my reflecting consciousness directed at my desire to write a good paper is part of the same reflective state as the lower-order desire itself. There is only one act of consciousness at the reflective level. A contemporary mind-brain materialist might reasonably say that the entire reflective state is one global brain state with some areas of the brain directed at others. Once again, however, Sartre struggles on the very same page with how to phrase the matter: 'It is agreed then that reflection must be united to that which is reflected-on by a bond of being, that the reflective [i.e. reflecting consciousness] must be the consciousness reflected-on. *But on the other hand*, there can be no question of a total identification of the reflective with that reflected-on' (BN 213, emphasis added; cf. BN 395-6). I believe that figure 4 best represents

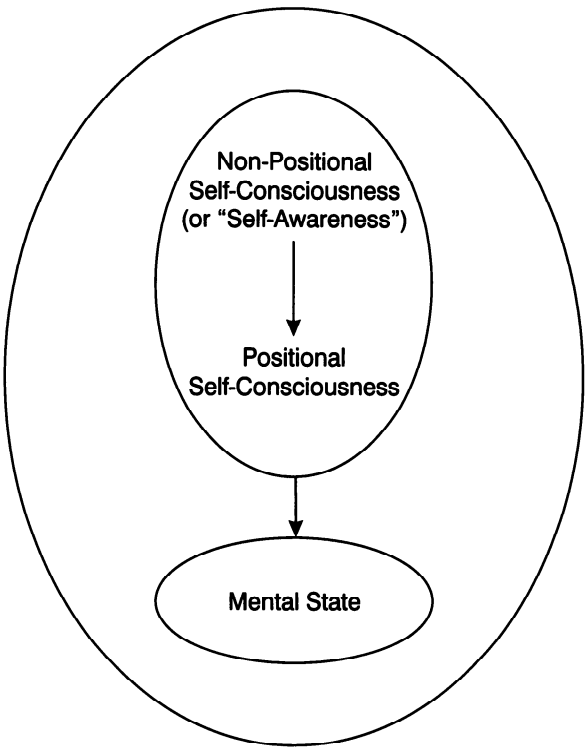

Figure 4.

what Sartre means here. There is a 'unity' and a 'bond' between the reflecting consciousness and the reflected-on consciousness, but they cannot be 'totally identified' with each other. Rather, as the above quotations make clear, they are part of one 'being' (i.e. one conscious mental state) 'fused' together into an 'indissoluble unity.'

In summary and restricting the matter to the reflective level, it might be helpful to put the unity problem by saying that Sartre obviously cannot consistently hold all of the following propositions:

(1) Knowledge requires a separation between the knower and the known object.

(2) Reflection involves knowledge of the 'reflected-on' state.

(3) The 'reflected-on' is not completely detached from the 'reflecting' state.

I have conceded that there is much confusion and ambiguity in Sartre's use of the terms 'knowledge,' 'separation,' and 'detachment.' If he means 'literal separation' implying two distinct 'objects' or mental states, then indeed we have an inconsistent triad; that is, he could not hold that both (1) and (3) are true if we accept his initial adherence to (2). Of course, if we allow Sartre to back off of (2), then (1) and (3) remain consistent but at the cost of trying to understand the difference between 'recognition' and 'knowledge.' However, Sartre could mean 'virtual separation' in (1) which I suggest could yield the more coherent position represented in figure 4. Recall the quotation from Sartre that 'presence to always implies duality, at least a *virtual separation*. The presence of being to itself implies a detachment on the part of being in relation to itself.... But if we ask ourselves ... *what it is* which separates the subject from himself, we are forced to admit that it is *nothing*' (BN 124, first emphasis added). All three claims could then be consistent. Proposition (1) is true because a virtual separation can allow for knowledge *within* the very complex reflective state. Proposition (2) is true since the reflecting state would have knowledge of the reflected-on state. And (3) could also be true because, as figure 4 indicates, the reflecting state and the reflected-on state are not completely detached but are part of the same complex reflective state. We can then also understand what Sartre meant by a 'quasi-object' when he said that 'the reflected-on is not wholly an object but a *quasi-object* for [pure] reflection' (BN 218, emphasis added). A quasi-object, in this context, would be a reflected-on mental state that is only *virtually* separated from the reflecting state. A quasi-object is not an entirely *distinct* object of knowledge.[53]

Unfortunately, one further complication arises when Sartre says such things as:

> reflection — if it is to be *apodictic [i.e. certain] evidence* — demands that the reflective *be* that which is reflected-on. But to the extent that reflection is *knowledge*, the

---

53 As we saw in Section III.1, it is perhaps once again open to Rosenthal to argue that, even though the HOTs are distinct and extrinsic from their targets, the unity in question is the amalgam of both states. The problem for Rosenthal here is twofold: (a) As was mentioned in Section III.1, if 'unity' means 'one single conscious state,' as Sartre ultimately seems to believe, then much of what Rosenthal says is at odds with this way of understanding the HOT theory. (b) On the *reflective* level, recall that Rosenthal's theory has three distinct states in contrast to what is expressed in figure 4. This would then make it even more difficult for him to treat the unity in question simply as an amalgam of three distinct parts. As I hope I have made clear in this section, however, I believe that we are all struggling to make sense of this 'parts in a unity' idea. Nonetheless, I have argued that the WIV holds out the best prospect for success.

> reflected-on must necessarily be the *object* for the reflective; and this implies a separation of being. Thus it is necessary that the reflective simultaneously be and not be the reflected-on. (BN 213-14, first emphasis added)

Our analysis thus far can handle the last part of this quotation, but another aspect of the unity problem results from Sartre's apparent desire to maintain a Cartesian infallibility thesis about one's own mental states. Indeed, Wider prominently mentions this thesis at the very beginning of TLG. Any claim of infallibility, of course, goes well beyond any ordinary notion of knowledge that we have thus far considered. But, just as we saw in Section III.1 regarding Sartre's related rejection of the unconscious, things are not so simple. In TE we find this rather striking passage:

> Is it therefore necessary to conclude that the [reflective] state is immanent and certain? Surely not. We must not make of reflection a mysterious and infallible power, nor believe that everything reflection attains is indubitable *because* attained by reflection. Reflection has limits, both limits of validity and limits in fact. (TE 61-2; cf. TE 65)

So despite Sartre's frequent talk of the 'translucency' of consciousness, it is not at all clear that some kind of Cartesian 'transparency' or 'infallibility' follows. Many others also have serious doubts about Sartre's adherence to such a strong Cartesian position. Morris questions any inference from 'translucency' to 'transparency' and tells us that 'Sartre neither says nor means that consciousness is always perfectly transparent.'[54] Catalano argues against the inference from translucency to infallibility: 'Sartre's claim that consciousness is translucent does not imply that we always have a correct understanding of that which we are aware.'[55] And Brown and Hausman question Sartre's commitment to the claim that the mind is translucent to itself by disputing that Sartre believed in the transparency and infallibility of the mental.[56]

There is one final and very puzzling aspect of Sartre's view that is worth briefly raising at this point. Recall from Section I that Sartre says that consciousness or the for-itself 'is not what it is and is what it is not' (BN 120, 127). This is commonly taken as Sartre denying that the Law of Identity, which says that each thing is identical to itself, applies to

---

54 Phyllis Sutton Morris, 'Sartre on the Self-Deceiver's Translucent Consciousness,' 105

55 Joseph Catalano, 'Successfully Lying to Oneself: A Sartrean Perspective,' 680

56 Lee Brown and Alan Hausman, 'Mechanism, Intentionality, and the Unconscious: A Comparison of Sartre and Freud,' esp. 541ff. and 552

consciousness. Indeed, such a view plays a prominent role throughout BN. But how could this be? How could anything not be identical with itself? Now some of Sartre's discussion of this matter goes well beyond the scope of this paper and has to do with the *temporality* of consciousness, which is a major topic in its own right.[57] However, Sartre also speaks as if such a denial applies to the *structure* of conscious mental states, which is our primary concern in this paper. Moreover, he seemed to think that denying the Law followed from his view that consciousness is self-consciousness. I believe that Sartre was mistaken, but let us see exactly why. One place where he discusses this issue is back at BN 114 in describing conscious belief:

> To believe is to know that one believes, and to know that one believes is no longer to believe. Thus to believe is not to believe any longer because that is only to believe — this is the unity of one and the same non-thetic consciousness.... To believe is not-to-believe. (BN 114)

The idea seems to be that conscious mental states and, in this case, conscious belief is not identical to itself. As Wider puts it: '[Conscious belief] is and is not what it is. Belief is belief but because it is self-conscious it is not belief' (TLG 334). I agree with Wider when she says that she fails 'to see how it follows from the fact that self-awareness is a property of pre-reflective consciousness that an act of consciousness at the pre-reflective level, such as belief, is not itself' (TLG 336). But given our analysis in this and the previous section, we can more clearly see Sartre's confusion here. I have argued that pre-reflective conscious states, for Sartre, have a dual and complex nature including (in this case) *both* a world-directed belief *and* a non-positional self-awareness or self-consciousness (of) the belief. Let us call the former ('lower') part 'part 1' and the latter ('upper') part 'part 2,' and simply refer to the entire pre-reflective state as the 'whole.' Now the whole is surely identical with the whole, part 1 is surely still identical with part 1, and part 2 is still identical with part 2. The Law of Identity remains unthreatened. Of course, it is true that the whole is neither identical with part 1 nor with part 2, but that would not violate the Law of Identity. In a similar way, acknowledging that part 1 is not identical with part 2 does not violate the Law. On the WIV and on Sartre's view, we can see how confusion might arise. Nonetheless, conscious mental states are still identical with themselves. Clarifying the structure of conscious states as we have in the

---

57 I cannot pursue this topic here, but see BN (especially Part Two, Chapter Two entitled 'Temporality') and BNC 43-57 & 150-4.

last two sections and in figure 3 helps us to see what Sartre apparently did not. Thus, Sartre is wrong when he replies that 'to affirm that the consciousness (of) belief [= part 2] is consciousness (of) belief is to dissociate consciousness from belief, to suppress the parenthesis, and to make belief an object for consciousness' (BN 121). But part 2 can clearly be identical with itself *and* we can *also* maintain that consciousness is not 'dissociated' from the belief while keeping the parenthesis and not making belief a distinct object of consciousness.[58]

V The BN 11 Argument

In chapter four of BNC, Wider considers several alleged counterexamples to the general thesis that consciousness entails self-consciousness (hereafter, the CESC Thesis) and discusses how Sartre might reply.[59] In that chapter, however, Wider also discusses what I consider to be Sartre's main argument for his belief that all consciousness is self-consciousness. As I have made clear throughout this paper, I agree with many of Sartre's conclusions about the structure of conscious states as well as being a defender of the CESC Thesis. However, this is not to say that I agree with all of his reasoning for that thesis, and his main argument for it is shaky at best. This often quoted passage goes as follows:

> the necessary and sufficient condition for a knowing consciousness to be knowledge *of* its object, is that it be consciousness of itself as being that knowledge. That is a necessary condition, for if my consciousness were not consciousness of being consciousness of the table, it would then be consciousness of the table without consciousness of being so. In other words, it would be consciousness ignorant of itself, an unconscious — which is absurd. This is a sufficient condition, for my being conscious of being conscious of that table suffices in fact for me to be conscious of it. (BN 11)

What is going on here? Well, the claim of sufficiency seems fine, but what about the claim of a necessary condition? That is, why exactly is such

---

58 And, of course, the mere fact that mental states are directed 'outside of themselves' does not violate the Law either. To think so would be to confuse the mental state with *the content* of the mental state.

59 They have to do with dreaming, people with blindsight, and Armstrong's well-known long-distance truck driver case. Obviously, I do not believe that such cases threaten the CESC Thesis, but I will not digress into a lengthy discussion of them here. It should be noted also that Wider defends Sartre and the Thesis to some extent later in BNC 164-169. She is right, however, that if Sartre completely rejected the unconscious, then it is much more difficult for him to handle the blindsight cases.

self-consciousness necessary for having a conscious mental state of, in this case, perceiving the table? Sartre uses a *reductio ad absurdum*; namely, that if there were no self-consciousness, then there would ultimately be 'consciousness ignorant of itself, an unconscious — which is absurd.' The idea seems to be that an unconscious consciousness is clearly absurd. However, while saying that *the same* mental state is both conscious and unconscious would be absurd, Sartre's *reductio* does not really force us into that conclusion. The reason is that he is referring to two different mental states (or at least two parts of a complex mental state): (1) the first-order world-directed consciousness of the table, and (2) the self-consciousness (of) the first-order consciousness. Sartre needs to show that (1) cannot occur without (2) and his main argument does not really do so, since an unconscious (2) does not contradict having a conscious (1). At the least, much more argument is needed to establish that (1) cannot exist without (2). So Sartre's argument at best begs the question, and, as Wider puts it, the question remains: 'is a consciousness that is not self-consciousness absurd, a contradiction?' (BNC 104). I believe that Sartre is attempting to give some support for the HOT theorist's intuitive idea, mentioned early in section II, that when one is in a conscious mental state one is certainly aware that one is in it. (Recall also that the sense of 'conscious state' at work is Nagel's sense, i.e. there is 'something it is like to be in that state' from a subjective or first-person point of view.) However, Sartre's BN 11 argument for the CESC thesis fails partly because he is not properly distinguishing between the lower-order and the meta-psychological state. Wider rightly points out (at BNC 104-5) that there are many other people who have argued for some version of the CESC thesis, including David Rosenthal and Roderick Chisholm, but then she is quick to mention many who disagree with it (including Fred Dretske). I do not wish to enter into this debate here, but Wider is right that it is incumbent on 'Sartre and others who hold such a view [to] offer a defense of this position' (BNC 105).[60]

Returning to the B11 argument, I think that it fails also because Sartre is not clear about the distinction between what Rosenthal calls 'transitive' and 'intransitive' consciousness.[61] Rosenthal explains that we sometimes use the word 'conscious' as in our being 'conscious of' something. This is the transitive use. On the other hand, we also apply the term 'conscious' to mental states, and 'the lack of a direct object suggests

60 Indeed, this is precisely what I have done at book length in CSC.

61 Rosenthal, 'A Theory of Consciousness'

calling this the intransitive use.'[62] Using this distinction, Rosenthal characterizes the HOT theory as claiming that 'a mental state is intransitively conscious just in case we are transitively conscious of it.' So, for example, my perception of the table is (intransitively) conscious just in case I am (transitively) conscious of it. Rosenthal makes it clear that analyzing intransitive (i.e., state) consciousness in terms of transitive consciousness is not circular because transitive consciousness is not a type of intransitive consciousness. Sartre is led astray because he conflates the two and therefore claims that there is a contradiction where there is not. But even if we agree that the B11 passage fails *as an argument* for the CESC thesis, some of it can still be interestingly viewed as a (perhaps unclear) version of Rosenthal's *statement of the thesis*. In defense of Sartre, he might generously be interpreted as stating that intransitive consciousness entails transitive consciousness even if he has not shown that the opposing view is absurd or contradictory.

But we must be careful here. Although Rosenthal addresses why his analysis is not circular, does the threat of an infinite regress reappear? As we saw in Section III.2, this is a concern of both Sartre and any HOT theorist. Although Rosenthal uses the term 'conscious of' in speaking of transitive consciousness, he immediately notes that he is *in that context* using 'interchangeably the notions of being conscious of and being aware of.'[63] So the HOT is aware of [i.e. transitively conscious of] the lower-order state, but presumably such 'awareness' is not conscious in the Nagelean sense; otherwise there would be the infinite regress that Rosenthal is so careful to avoid in his earlier paper.[64] This terminological matter is perhaps a bit misleading and confusing, and it is why I normally prefer to avoid the transitive/intransitive consciousness distinction. Nonetheless, Sartre seems to be attempting something similar to Rosenthal; namely, analyzing intransitive consciousness in terms of transitive consciousness. Indeed, as we have seen through the numerous quotations above, Sartre frequently uses the transitive 'conscious of' way of speaking throughout BN. Nonetheless, we must be careful not to interpret the transitive 'conscious of' locution as implying anything more than some kind of nonconscious awareness.[65]

---

62 Ibid., 26. The next quotation is also from 26.

63 Ibid., n. 28

64 Rosenthal, 'Two Concepts of Consciousness'

65 For another criticism of the BN 11 passage, see Soll, 'Sartre's Rejection of the Freudian Unconscious,' 593-6.

VI I-thoughts and I-concepts

Another contention of the HOT theorist is that the HOT (or MET) is of the form 'I am in mental state M now.' That is, the content of the HOT is an indexical thought making reference to both the person having the mental state (the 'I') and the mental state (the 'M'). However, it is crucial to recall that the HOT is not itself conscious when one is in a world-directed conscious state. The 'I-thought' is *implicit* in such cases and only becomes *explicit* on the reflective level. For example, my current conscious desire to work on this paper contains the implicit (unconscious) thought that I am having that desire now. But when I reflect or introspect, I consciously and explicitly apprehend that desire as mine. The same is true when we have all sorts of other mental states, e.g. beliefs and various perceptual states.

Sartre initially has difficulty with this aspect of the HOT theory while distinguishing between pre-reflective and reflective conscious states. In TE he first asks, 'is there room for an I in such [an unreflected] consciousness? The reply is clear: evidently not' (TE 41). But this is where, in TE, Sartre is concerned with rejecting Husserl's transcendental I. Thus I think that Sartre is merely saying that on the pre-reflective level there is no *conscious* apprehension of an I (i.e. no conscious HOT) because my conscious attention is focused outside of me, as the HOT theory also claims. So later Sartre explains that 'there is no I on the unreflected level. When I run after a streetcar, when I look at the time, when I am absorbed in contemplating a portrait, there is no I. There is the consciousness *of the streetcar-having-to-be-overtaken*, etc., and non-positional consciousness of consciousness' (TE 48-9). Once again, the HOT theorist would say that there is no *conscious* I on the unreflected level, and I believe that is what Sartre meant in this passage. There is, however, a nonconscious or implicit I even on the unreflected level. Sartre says as much later in TE: 'It is certain ... that the I does appear on the unreflected level' (TE 89). When I am preoccupied with outer reality (e.g. in trying to hang a picture or in running after a streetcar) I am still *implicitly* aware of myself ('I') as being in a conscious mental state. In BN Sartre explains that the 'unreflective consciousness is a consciousness *of* the world. Therefore for the unreflective consciousness the self exists on the level of objects in the world.... Only the reflective consciousness has the self *directly* for an object' (BN 349, second emphasis added). So a first-order consciousness is directed at the world and a reflective consciousness is directed at oneself, but, even on the unreflective level, one still has *indirect* or *implicit* thoughts about oneself.

Now it is clear that there are many kinds of so-called 'I-thoughts' ranging from very unsophisticated to very sophisticated. Such sophistication, in part, results from the type of 'I-concept' contained within the

I-thought.[66] So, for example, we might distinguish between the following I-concepts:

1. I *qua* thinker among other thinkers.
2. I *qua* enduring thinking being.
3. I *qua* experiencer of a current mental state.
4. I *qua* this thing as opposed to other physical things (where 'this thing' refers to one's body as distinct from other bodies).

The above I-concepts are in decreasing order of sophistication. Thinking of myself as a thinker among many other thinkers or as an enduring object through time requires conceptual abilities that go beyond simply distinguishing my body from other objects. Thus, the type 4 concept can presumably be possessed by various lower animals whereas the type 1 concept may only be a human or higher mammal capacity.[67] The point I wish to emphasize here with respect to Sartre is that, depending on the mental capacities of the organism, one can have either an unconscious or conscious HOT which contains any of the above I-concepts. Most relevant for my purposes is showing how this point relates to Wider's interpretation of Sartre as emphasizing what she calls 'bodily self-consciousness' (BNC 115). In chapter five of BNC, Wider develops an interpretation of Sartre whereby 'the most basic form of self-consciousness must be bodily awareness' (BNC 115). I am largely in agreement with this view both as an interpretation of Sartre and as an independently plausible theory. Indeed, Wider cites numerous examples of philosophers and psychologists, such as Gareth Evans, Gerald Edelman, and Maurice Merleau-Ponty, who also seem to hold such a view in some form or another (BNC 127ff. and 139ff.). Wider states the principle claim as 'the body is the subject of consciousness' (BNC 115). Thus, an implicit 'I' refers to the body even when one has the most rudimentary conscious state on the unreflective level. It seems to me that Wider's view can be interpreted as endorsing the notion that the type 4 I-concept above is the least sophisticated and that it appears even on the unreflective level. Wider explains:

> all consciousness, even at the prereflective level, must be present to itself. Now if consciousness just is the body's presence to the world, as Sartre argues, then the

---

66 Indeed, this is what I argue at greater length in CSC 78-84.

67 See again CSC 78-84, but also CSC ch. 9. Actually, I argue that any conscious lower animal is at least capable of having both the type 2 and the type 3 I-concept.

> body must be present to itself in being present to the world. So there must be a kind of consciousness of the body, what I will call bodily self-consciousness, and this must form part of our awareness of the world. The most basic form of self-consciousness must be bodily awareness. (BNC 115)

So, in terms of the HOT theory, even when we have first-order world-directed conscious states, there is bodily self-consciousness; that is, a HOT (or MET) containing at least the most primitive (type 4) I-concept. Such an I-concept is thus part of the I-thought which, as I have argued in previous sections, is part of the complex conscious state itself. So, once again, all consciousness is self-consciousness. That such bodily self-awareness even exists on the unreflective level is indicated by Sartre's remark that 'the body belongs ... to the structures of the non-thetic [i.e. non-positional] self-consciousness' (BN 434).[68]

Much of this is reminiscent of the Kantian idea that having experience of outer objects presupposes distinguishing them from oneself. Thus having I-thoughts is presupposed in objective experience. At the very least, in order to having conscious thoughts about external objects one must be able to differentiate them from oneself (including one's own body and one's own mental states). If one did not implicitly distinguish outer objects from oneself, then one would treat the enduring objects of experience as merely momentary fleeting subjective states which, in turn, would make objective experience impossible. Restricting ourselves to bodily self-awareness, then, Wider tells us in a Kantian spirit that 'without bodily self-consciousness, consciousness of the world is impossible' (BNC 118). Citing such prominent philosophers as Daniel Dennett and Owen Flanagan, Wider explains that 'even on the most primitive levels of conscious life, simply to survive an organism must be able to distinguish its biological self from that which is not itself. There must be an ability to make a me/not-me distinction' (BNC 122-3). As I mentioned above, it would seem likely that even the lowest of conscious creatures are at least capable of having the type 4 I-concept. This would therefore support the idea that even the most primitive conscious creatures are self-conscious in at least this rudimentary way.[69]

---

68 It occurs to me that the above four I-concepts correspond somewhat to Sartre's distinction between four types of self-consciousness (see BNC ch. 3) in the following way: (a) type 1 = being-for-others; (b) type 2 = impure reflection; (c) type 3 = pure reflection; and (d) type 4 = pre-reflective (or 'bodily') self-consciousness.

69 For more on the connection between Kant and the HOT theory, see CSC chs. 3, 4 and 9. For more on the connection between Sartre and Kant, see BNC 20-39 and TE 31ff.

I must confess, however, that I do not see how any of the above or anything in Wider's book warrants her blanket rejection of identifying conscious mental states with neural activity. In a very puzzling passage, Wider says that many philosophers, such as John Searle and Colin McGinn, are guilty of ignoring the rest of the body when they speak of consciousness as a brain property. Wider says that 'it is the body, the organism as a whole, that *is* conscious' (BNC 114, emphasis added). It is not clear to me what this means other than what she says two sentences later: 'the body (not the brain) ... *is the subject of* consciousness' (BNC 114, emphasis added). But even with our explication above of this latter claim, it clearly does not follow that the entire body *is* conscious. Rather, one has a kind of self-awareness which *refers to* one's own body (an 'I-thought'). But the complex conscious mental state *itself* can still be identified with neural activity, though the HOT theorist is typically silent on this empirical question.

Wider also seems to be confusing what Rosenthal calls 'creature consciousness' with 'state consciousness.'[70] The former recognizes that we often speak of whole organisms as conscious, but the latter recognizes that we also speak of specific mental states as being conscious. Indeed, it is state consciousness that has been my primary concern throughout this paper and Sartre is also clearly first and foremost attempting to present a theory of state consciousness. Thus when one is concerned with examining what makes particular mental states conscious, a mind-brain identity theory can still arguably be the most plausible reductionist alternative while also remaining compatible with everything said earlier in this section. In short, conscious mental states can still *be* neural states while also *referring to* the body and having the structure defended throughout this paper. Bodily aware I-thoughts accompany every conscious mental state, but it does not follow that the entire body *is* conscious. My conscious visual perception of the tree contains implicit reference to my body, but the visual experience itself can still just be identical with a pattern of neural activity in my visual cortex. Indeed, it would not even seem to make sense to say that such a conscious mental *state* is my body.

Perhaps even more puzzling is Wider's analogy that 'just as a jet engine needs wings and tail and other parts to generate flight in the aircraft, so too the brain needs kidneys and lungs and blood and air to generate consciousness' (BNC 114). If this is meant to support the identity of the body and consciousness, then it also fails. First of all, in the

---

70 Rosenthal, 'A Theory of Consciousness'

case of an airplane there is no analog to state consciousness; that is, we only speak of the entire airplane flying and never just of an engine flying. Second, Wider is confusing a necessary condition with an identity claim: it is one thing to say that A *needs* {B, C, ... X} *in order to generate* Z, but quite another to say that {A ... X} *is* Z. Once again, the latter does not follow from the former. It is still perfectly possible that only A (or a part of A) *is* conscious. Even Wider seems to concede this in her analogy by saying that the 'brain ... generate[s] consciousness' (albeit with the help of the necessary conditions she mentions, such as blood and air). Thus we must be careful not to take Sartre's theory as ruling out a mind-brain identity theory.

VII Conclusion

Although Sartre may not have always argued well for his conclusion that consciousness is self-consciousness and although he did often struggle with the details of his theory (such as with the unity problem), I hope I have shown that he had much of value to say about the structure of conscious mental states. For example, Sartre's theory is importantly related to the HOT theory of consciousness and he held a modified version of the theory. I have therefore shown how his views can be informatively placed against the background of a contemporary analytic approach to consciousness. Sartre also addressed several key issues frequently associated with the HOT theory, such as the threat of an infinite regress and the debate over the existence of nonconscious mentality. Finally, as Kathleen Wider has helped us to understand, Sartre's theory also offers insights into the important relationship between having conscious mental states and the presence of so-called 'I-thoughts.'[71]

Received: May, 2001
Revised: August, 2001
Revised: March, 2002

---

71 Thanks to Kathleen Wider and Yiwei Zheng for some helpful correspondence during my work on this paper. Thanks also to a referee for several helpful comments.