



Is Self-Representation Necessary for Consciousness?

William Seager
University of Toronto at Scarborough
Humanities Building Room 523
215 Huron St.
Toronto, Ontario
Canada M5S 1A2
© William Seager
seager@utsc.utoronto.ca

PSYCHE 12 (2), May 2006

KEYWORDS: Consciousness, Self-Representation, HOT Theory, Introspection

Abstract: Brook and Raymont do not assert that self-representing representations are sufficient to generate consciousness, but they do assert that they are necessary, at least in the sense that self-representation provides the most plausible mechanism for generating conscious mental states. I argue that a first-order approach to consciousness is equally capable of accounting for the putative features of consciousness which are supposed to favor the self-representational account. If nothing is gained the simplicity of the first-order theory counts in its favor. I also advance a speculative proposal that we are never aware of any distinctively mental attributes of our own states of consciousness except via an independent act of reflective conceptualization, although this goes rather farther than the first-order theory strictly requires.

In their provocative and highly compressed discussion, Brook and Raymont (henceforth BR) raise a host of key issues about what they call the representational base of consciousness. The heart of their proposal involves two ideas: 'self-representing representations' and 'global representations'. Space limitations prevent me from delving into more than a few of the fascinating points BR discuss. I will focus on self-representation and in particular on the question of whether any features of consciousness force us to postulate that conscious states are self-representing.

What are self-representing representations? I assume that every representation can be characterized in terms of its content and its vehicle. The sentence 'cats are evil' has as its content the proposition that cats are evil and its vehicle, in this case, is the black marks on my computer monitor. A self-representing representation is one that includes reference to itself in its *content*. So, the sentence 'this is a sentence' is a self-representing

representation. Consider this set of sentences {'there is a sentence in this set', 'cats are evil', 'there is a false sentence in this set'}. This is not a case of self-representation because none of the representations refer to themselves (the set is not itself a representation). Thus, it would not follow from the fact that the mind is a set of representations which includes representations about that set, that the mind contained self-representing representations. Many mental states, if not all, are representational states and it is very likely that their vehicles are brain states. The bare fact that some of these states are about brain states would not make them self-representing states. Representations can be combined to form more complex representations but self-representation does not arise just from conjoining two representations, R1 and R2, into [R1 & R2] where R2 is about R1. At least, cases where two discrete representational components of a complex representation are such that one component is about the other component are not interesting forms of self-representation in the context of the issue of whether conscious states are necessarily self-representing since the self-representation in such cases can, so to speak, be disentangled. We might label the interesting form of self-representation that conscious states are supposed to enjoy, 'essential' self-representation. I will argue here that BR do not show that essential self-representation is necessary for consciousness.

There is a classic argument in favor of the idea that consciousness involves essential self-representation that can perhaps be traced back to Aristotle (see Caston 2002) and was explicitly advanced by Brentano (1874/1973). The argument assumes that all mental states are conscious and that all conscious states are 'self-intimating' in the sense that any creature who is in a conscious mental state is aware of that state. Unless at least some conscious states are self-representing in the sense that they do not require an additional state whose direction onto the original state underwrites that state's self-intimating nature, an infinite and vicious regress threatens. For, supposing that self-intimation is dependent upon a further mental representation (presumably one that targets the original state as its 'intentional object'), then that additional state will (by the first assumption) be itself a conscious state and will then, by the second assumption plus our current assumption about the source of self-intimation, require a third mental state, and so on to infinity. And this regress is vicious since no state can be conscious unless it is self-intimating (it is assumed) and this requires the completion of the infinite sequence of mental states.

However, this argument is far from convincing. I think it is a fundamental mistake to accept that conscious states are self-intimating: it is the mistake of confusing the message with the messenger. There is no need for us to be aware of states of consciousness in order to be fully conscious of the contents of those states. I would go so far as to advance a principle:

(P) we are never aware of any distinctively mental attributes of our own states of consciousness (e. g. that they are conscious states) except via an independent act of reflective conceptualization.

Of course, conscious experience is available to be conceptualized *as* experience, but the principle entails that there is *never* any direct awareness of states of consciousness as such. One must always take up a reflective stance in which one regards what is happening as conscious experience. However, this – perhaps rather extreme – point of view cannot

be defended here, although it will unavoidably intrude into the discussion below.

In any case, the regress argument suffers from a less contentious and oft recognized flaw: unconscious mental states. The existence of unconscious mental states is no longer at all controversial. Typical HOT theories of consciousness (see Rosenthal 1986, Carruthers 2000) exploit their possibility to halt the regress. While the existence of unconscious mental states destroys the regress argument, it also puts pressure on the idea of self-intimation. While at first glance it seems intuitively obvious that an unconscious mental state is not self-intimating, one could consistently maintain that unconscious mental states are *unconsciously* self-intimating. If, for example, self-intimation is explicated in terms of immediate, non-inferential knowledge (or perhaps belief), so that a state, S, is self-intimating just in case x's being in S implies that x knows (believes) that he is in S, then one could hold that there could be unconscious knowledge (belief) of being in S. In fact, the idea of unconscious knowledge (or unconscious beliefs) is not strange and in fact is the most typical example of unconscious mental states.

Somewhat curiously, the HOT theories' method of halting the regress makes it impossible for a mental state to be unconsciously self-intimating, since according to HOT theory, if being in S immediately and non-inferentially brings about the belief that one is in S then S is a conscious state. There seems to be no reason to deny that unconscious self-intimation could be immediate and non-inferential—in fact, given its nature this is the most natural way to view unconscious self-intimation. Appeal by HOT theorists to some unspecified notion of special kinds of higher-order thoughts which are 'suitable' to generate consciousness is blatantly *ad hoc* given the initial, and essential, acceptance that all kinds of mental states can occur unconsciously. So unconscious mental states cannot be self-intimating. This naturally raises the question of why self-intimation has this power to make a state into a conscious state and exactly why it is impossible to have unconscious knowledge of being in an unconscious mental state.¹

As BR note, several lines of objections against HOT theory exploit this feature: that S's self-intimation works via S producing a higher order state about it. I tend to think that BR are right to reject HOT theories on the basis of these objections and won't review them here.²

If by 'self-intimating' is meant that one cannot be in a conscious state, S, without *knowing* (or at least believing) that one is in S then it seems highly plausible that there are non-self-intimating mental states. In order to know that one is in S one needs the concept of S and the huge majority of conscious beings are utterly bereft of such concepts (this is of course the basis of further objections against HOT theories which have frequently been made; see e. g. Seager 2004).

In line with this, BR wish to separate the notion of self-intimation from self-representation. What they say is

We are not saying that having a representation guarantees being conscious of it and oneself. The view we are advancing is only that having a representation is everything representational needed to become conscious of having it and of oneself as its subject. Other things may be needed, too: the ability to direct attention onto one's own psychological states, for example, or the conceptual resources to go from merely representing something, say a computer, to representing it as a computer.

The self-representing feature of conscious states thus does not *suffice* to make a subject aware of those states. Is self-representation *necessary* for becoming aware of one's conscious states? I have my doubts. One way in which BR argue that it is necessary depends upon mental states having features of their own of which we are or can be aware when we are conscious of those states. For example, BR claim that when we are conscious of something we are also aware of *how* it is being represented. They say: 'some aspects of how an object is represented (visually, aurally, etc.) seem to be aspects of the representation itself'.

It is important to be careful about the phrase 'aspects of the representation itself'. Consider the word 'mermaid'. If asked to provide information about that 'representation itself' I would most naturally take this question to be about the vehicle of the representation. I would thus note that it is a *word* of English realized, here, as black marks upon a computer monitor. In the case of mental representations, physicalists, at least, would characterize mental representations in terms of neurological attributes. This cannot be what BR intend, for we are certainly never aware of these aspects of our mental representations. I take it that it is not aspects of the representation itself that we are aware of according to BR but rather aspects of the *experience* itself, apart from what it reveals about the 'external world'.

So what are the 'aspects' of the experience itself that we *are* aware of? Obviously it is not perceived colour, shape or motion, for these are features assigned to the objects of perception rather than features of the experience itself. It is not at all clear what these 'aspects' could be. What is it about a visual experience that 'reveals' that it is *visual* rather than, say, *auditory* except the features of the environment which are presented in consciousness—colours and shapes for vision, pitch and timbre for hearing? Nothing could be more dissimilar than colours and sounds, and this difference seems sufficient to ground our awareness of the difference between seeing and hearing.

But suppose, for the sake of the argument, we grant that there are features of the *experience* itself of which we can be aware and which are not features of the object of awareness as ordinarily conceived. The problem is that it seems that this claim is compatible with there being a representational mechanism of the same sort that represents external objects but which in this case represents features of mental states. The linchpin principle of representational accounts of consciousness is that in consciousness we are aware only of the *contents* of certain representational mental states.

The point here is that nothing prevents there being representations of (features of) mental states, and this permits the representationalist to accept awareness of mentalistic features without invoking self-representation. The mechanism of our consciousness of such features would then be just the usual representational machinery, aimed at the 'internal' rather than 'external' world (so to speak). There is no special need for self-representing representations to make this possible.

Such an account does not imply that a state's being conscious requires that it be the target of a higher order representation. At most, the consciousness of certain distinctive features of mental representations would require a representation of them. Some states target the colour of birds, some the putative 'qualitative' aspect of states of consciousness. It is not in virtue of the representing state representing a mental state that

the latter is conscious.

BR raise the general worry against HOR approaches, that the existence of the higher-order representing state will be sufficient all by itself to generate a state of consciousness. This leads to a difficult dilemma for HOR theories: either there can be ‘hallucinations’ of consciousness, wherein it seems to the subject that they are conscious when they are not, or the higher order state is sufficient to generate consciousness and the relational structure of HOR accounts is otiose. Neither horn is attractive to the HOR theorist, but no such dilemma faces the FOR account of awareness of mentalistic features sketched here. FOR accounts equate states of consciousness with (certain) representational states. These representational states can, of course, be more or less accurate, but that is irrelevant to whether they are conscious or not. Imagine, then, a case where the postulated representational state which represents a distinctively mentalistic feature that typically figures in some sort of conscious state, say the ‘visualness’ of conscious visual perception, occurs when the subject is not engaging in visual perception of any kind. Let us suppose that the subject is merely dreaming that he is watching a school of mermaids at play around a sunken wreck. The FOR analysis of such a state of consciousness, under the assumption that there are mentalistic features of visual experience in addition to the host of ‘external’ features such as colour, shape, position etc., is that there is a representation of swimming mermaids plus a representation of the ‘visualness’ of the former representation. We are assuming that this is a conscious state, and the content of that conscious state is given by the content of the representations at issue. This content includes, in addition to the colours, shapes, motions etc. of the mermaids, the mentalistic feature of ‘visualness’. Could this be hallucinatory, in the sense that this mentalistic feature is not present but merely ‘seems’ to be? It can certainly be hallucinatory in the sense that it represents a visual experience when no such experience is occurring; the dreaming subject thinks he is *seeing* when he is not. But it is not hallucinatory if we regard it from the point of view of how things are being represented: they are *really* being represented as ‘visually experienced’. Of course, this is exactly analogous to the mermaids: they are hallucinatory from the (normal) perspective of the issue of whether there are any mermaids hereabouts, but they are not hallucinatory in the sense that it is completely correct to say that the experience is ‘as of mermaids’.

Nonetheless, BR might urge that this account entails that it is possible for the representation of ‘visualness’ to occur in the absence of any other (conscious) representation. Such an experience is not easily describable; it would be an awareness of the visual aspect of perception in the absence of any external content which appears to be seen. But we know that severe dissociations are possible. There are, for example, patients who are blind to shape and position but yet can experience colour.³ Such dissociations are possible because the relevant features can be represented independently of each other. There is no a priori reason to rule out the possibility of dissociation of the representation of mentalistic features from that of non-mentalistic features (assuming the former to exist).

Even if we were to refuse to allow the possibility of such extreme dissociation, self-representation is not, by itself, going to forge a necessary link between the possession of ‘visualness’ and certain visually perceivable features of external objects. Like any representation, self-representation can be accurate or inaccurate. ‘This is not a sentence’

is a self-representing representation which misrepresents itself. Why could not a visual experience include a self-representational tag to the effect that ‘I am an auditory experience’? So if there is a problem here for the FOR account of awareness of mentalistic features of experience it is equally a problem for the self-representational approach.

There is a somewhat subtle feature of the point I am making here that deserves emphasis. The kind of awareness of mentalistic properties of experience I am exploring is the same kind of ‘direct’ awareness which FOR theories posit for typical aspects of experience such as colour and shape. I am positing that the FOR mechanism could be directed at mentalistic properties. At this point, there is no invocation of the idea that we have to conceptualize experience as experience in order to become aware of the ‘mentalness’ of experience.

Despite the foregoing, I want to stress that the admission that there are purely mental features of experience of which we are directly aware in consciousness is nothing but a concession made for the sake of the argument. It is a striking fact that we do not find cases of people misidentifying the mentalistic features of their experiences – being aware of visual experience as auditory, whereas we do find profound dissociations amongst the various external features represented in conscious experience. A natural explanation stems from principle (P) enunciated above. Awareness of the mental as mental is, according to (P), a reflective, conceptual task, and the idea that someone could reflectively introspect visual experience as auditory is about as plausible as someone reflectively miscategorizing the number 1 as an orang-utan.

BR make, rather in passing, another criticism of the FOR account when it is coupled to the account of our awareness of states of consciousness enunciated (in a very strong form) in principle (P). They claim that this makes our awareness of mentalistic features ‘inferential’. Therefore, it is important to note that (P) does not require endorsement of Dretske’s (1995) idea that the conceptualization of experience as experience is inferential. Sometimes concepts are applied via inference, but more frequently they are not.⁴ Typically, when I see a telephone *as* a telephone, I do not infer that there is a telephone in front of me – the concept is applied automatically, as it were. Similarly, when as a practised introspectionist, I attend to my own mental states, the relevant concepts are, in general, effortlessly or automatically applied to the whole range of conscious experience: visual, auditory, emotional etc. However, it is significant that sometimes I do have to rely on inferential mechanisms to become aware of my own mental state, just as sometimes I have to infer that the thing in front of me is a telephone.

My own view is that there is no need to posit any purely mentalistic features of experience and that a ‘pure’ FOR approach captures everything in consciousness. The particular examples which BR raise against FOR are controversial and have been given reasonably plausible representationalist interpretations by various FOR theorists which I need not review here. The present point is only to argue that the postulation of such features does not require us to embrace self-representing representations.

Neither the classic regress argument nor BR’s argument from putative mentalistic features of experience forces us to accept that mental states are self-representing. Of course, that does not show that conscious states are *not* self-representing, but it does

mean we need independent reasons for accepting self-representing states as the representational basis of consciousness.

Perhaps the motivation behind BR's drive towards self-representing states can be clarified by focusing on their remark that 'for at least a great many representations, simply having a representation is all the representing that one needs to do to become conscious not only of what is being represented but also of the representation itself'. As noted above, there is some unclarity about what it is to be conscious of a representation itself. This may mean: to be conscious of features of the representational *vehicle*. But what are the features of the representational vehicle? The representation is, let us suppose, some neural state and its own features will be some set of physical properties. But nobody is ever conscious of these (except maybe indirectly when hooked up to an auto-cerebroscope). Yet BR explicitly contrast what is represented with 'the representation itself'. But a self-representing representation represents in some way, or in some measure, its own features, so what the representation presents about itself is still 'what it represents'. Everything here is within the realm of represented content, and, as argued above, no special need for self-representation appears. Whatever the 'mentalist features' of the experience are, they have to be represented for them to be presented in consciousness. We might say that the system can either self-represent or other-represent these features, but on either view it is still necessary for these features to be represented.

There is another way to interpret the passage quoted above. It can be interpreted to mean that having the representation is all the representing that is needed to be conscious of *representing as such*. However, it is no more plausible that we are directly conscious of representing than that we are conscious of the vehicle of representation. We are only aware of what is *represented* (of course, it is possible to represent acts of representing so we can become aware of representing, but this requires additional representation). After all, many philosophers deny that the representational analysis of consciousness is correct, and while they are wrong, they are not refuted merely by appeal to introspection.

There is also a curious tension between the quoted remark and the carefully made distinction between what representing is needed for consciousness and what one is actually conscious of. BR say that 'other things' (beyond representation) will be needed for one to become actually conscious of one's mental states. They say that one might need 'the ability to direct attention onto one's own psychological states, for example, or the conceptual resources to go from merely representing something, say a computer, to representing it as a computer'. Why is representing something *as* a computer not additional representation? There are distinct truth conditions involved between just representing something which is a computer and representing it as a computer, so there would seem to be differences of content and hence a difference in representation. What would it be to 'direct attention to one's own psychological states' in a way that would make one become conscious of the state in question but of which one was not conscious before the act of attention? It would seem to involve coming to realize consciously that one was in a psychological state. This would seem to involve additional representational resources that were not in play before.

However, I think that BR are correct, if their claim is read properly, and they make an important point. The proper view is that a state, S, need be no more than

conscious for one to become aware of S as a state of consciousness. Since S's being conscious involves S presenting its content to the subject, all that is required for the subject to become aware of S as a conscious mental state is the kind of reflective conceptual activity enunciated in principle (P). No more 'representing' by S is necessary. And the conceptualization of S as a conscious state is a trivial task once one is in possession of the requisite concepts of mental states. But there is no need for S to be self-representing for one to take up this conceptual stance towards it.⁵

References

- Brentano, C. (1874/1973). *Psychology from an Empirical Standpoint*, London: Routledge and Kegan Paul.
- Byrne, A. (1997). Some like it HOT: Consciousness and higher-order thoughts, *Philosophical Studies*, 86, pp. 103-29.
- Carruthers, P. (2000). *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.
- Caston, V. (2002). Aristotle on Consciousness, *Mind*, 111 (444), pp. 751-815.
- Dretske, F. (1995). *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Milner, A. et. al. (1991). Perception and action in "visual form agnosia" *Brain*, Feb;114 (Pt 1B), pp. 405-28.
- Rosenthal, D. (1986). Two Concepts of Consciousness, *Philosophical Studies*, 49, pp. 329-59.
- Seager, W. (2004). A Cold Look at HOT Theory, in R. Gennaro (ed.) *Higher-Order Theories of Consciousness: An Anthology*, Philadelphia: John Benjamins.

Notes

1. It is not difficult to outline possible examples. One might be in some more or less complex and unconscious emotional state and give some signs that one is unconsciously aware of that state. Some of Freud's cases look something like this: an unconscious intention leads to actions which suggest an unconscious awareness of that intention. The point, of course, is not that Freud is correct but that the scenario is perfectly comprehensible.
2. Another line of objection is closely related to this one. It questions why self-intimation makes mental states conscious but not other states which can produce mental states which are about them (see Dretske 1995, Byrne 1997).
3. See for example Milner et. al. (1991) in which a patient is described who 'seems to have no awareness of shape primitives through Gestalt grouping by similarity, continuity or symmetry' despite the fact that 'visual acuity and colour vision, along with tactile recognition and intelligence, were largely preserved'.
4. It is impossible for all concept application to be inferential since the premises of the

inference must be judgments and hence involve conceptualization themselves. The vicious regress that threatens is obvious.

5. My thanks to David Bourget and Uriah Kriegel for helpful comments.