

The problem of higher-order misrepresentation

Graham Peebles

To cite this article: Graham Peebles (2021): The problem of higher-order misrepresentation, *Philosophical Psychology*, DOI: [10.1080/09515089.2021.2016677](https://doi.org/10.1080/09515089.2021.2016677)

To link to this article: <https://doi.org/10.1080/09515089.2021.2016677>



Published online: 21 Dec 2021.



Submit your article to this journal [↗](#)



Article views: 36



View related articles [↗](#)



View Crossmark data [↗](#)



The problem of higher-order misrepresentation

Graham Peebles

Intelligent Robotics Laboratory Graduate School of Engineering Science Osaka University Japan

ABSTRACT

The problem of higher-order misrepresentation poses a dilemma for the higher-order theory of consciousness. The two ways of conceiving of the theory each run into a different difficulty raised by the problem of misrepresentation. If the theory is conceived relationally, i.e., conceived so as the higher-order state causes or makes a first-order state conscious, then the theory faces a problem raised by Block concerning the implausibility of non-existent conscious states. If conceived non-relationally, i.e., conceived in such a way as it is the higher-order state itself which is the conscious state, then consideration of Neander's original misrepresentation problem, that this seems to make the first-order state a spinning wheel in the explanation of consciousness, raises the conceptual problem that this version of the higher-order theory seems not to offer a philosophical theory of consciousness that is distinct from the first-order theory.

ARTICLE HISTORY

Received 5 February 2020
Accepted 6 December 2021

KEYWORDS

Higher-order theory of consciousness; misrepresentation; targetless states

1. Introduction

The higher-order theory of consciousness is a type of functional-representationalism (see Brown, 2015). To use Block's (1995) canonical terminology, such a theory analyzes conscious states in terms of access consciousness: the representational content of those states being available for the immediate rational control of thought, speech, and action. This contrasts with the notion of phenomenal consciousness that characterizes conscious states in nonfunctional terms – “what it is like” to be in those states. According to the functional-representational theory, which comes in three broad categories, phenomenal consciousness can be analyzed as a type of access consciousness. The first-order theory analyzes conscious states in terms of their first-order content and their functional role. Thus, the sufficient condition for such a state M to be conscious is that it has the required functional role and first-order content, p .¹ In contrast, the higher-order theory requires that the first-order state M itself be content of a further, on pain of regress unconscious, state N which represents (or

“targets”) M and in virtue of which the subject is (unconsciously) aware of being in M and thus consciously aware that p .² The third category is the reflexive theory on which there is just the one state M which has two contents, both the usual first-order content p and also the second-order content of M itself, both of which the subject is consciously aware of.³

There are significant advantages that have attracted philosophers to the functional-representational (from now on just “representational”) theory. It analyzes consciousness in terms that are amenable to a further analysis in physical terms. Relatedly, and central to the subject of this paper, it offers a solution to the familiar conundrum of the objects of illusion and hallucination. As I write this, I am looking out of my office window at a plant on the terrace. But what if there is in fact no plant and I am hallucinating? In such cases, what is the object of experience? We tend not to pay as much heed, though, to what is essentially the same problem as it arises in non-perceptual cases, such as false belief. The proponent of the representational theory treats perceptual cases in the same way as non-perceptual cases. Hallucination is therefore handled easily: the state misrepresents how things are as it has a content, which is false/non-veridical, e.g., *that there is a plant on the terrace*, and so the putative object of the state, e.g., the “plant”, does not exist. It is this metaphysically innocent explanation of misrepresentation that provides the general theory with its powerful explanatory advantage over rival theories of consciousness.

But what happens on the higher-order theory if such misrepresentation occurs not at the first-order level but at the higher-order level? What if the higher-order state misrepresents the first-order state? The higher-order story of the veridical case of my seeing the plant, for example, is that I am in a first-order state with the content *that there is a plant on the terrace*, and a higher-order state with this state as its content – i.e., *I am having a visual experience of the plant*. In virtue of this higher-order state that makes me (unconsciously) aware of my first-order experience, I am in the conscious state of seeing the plant on the terrace. But what if the content of the higher-order state were to be not that the first-order state represents a plant on the terrace but a cat on the terrace? Or what if I were to be in a higher-order state that represents a first-order state that does not exist at all? Versions of this problem have been influentially raised against the higher-order theory (Byrne, 1997; Kriegel, 2009; Levine, 2001; Neander, 1998). Block (2011a, 2011b) raised it again, offering a version of the argument that has drawn responses from Rosenthal (2011) and Weisberg (2011) as well as discussion from others (Berger, 2014; Brown, 2015; Lau & Brown, 2019).

There are two conceptions of the higher-order theory and two versions of the problem of misrepresentation. Brown (2015) calls the two conceptions of the higher-order theory the “relational” and “non-relational” conceptions. Either the first-order state is an independent state and the targeting

relation essentially causes the first-order state to become conscious – to use the common metaphor, as if it were lit by a beam of light – or the two states are not really independent and the targeting relation, and thus the first-order state, is, so to speak, “internal” to the higher-order state (for unambiguous examples of both conceptions, see Wilberg, 2010; Gennaro, 2012 respectively). One could draw an analogy here with two ways of conceiving of the notions of intentionality and an intentional object. The first conception separates the first-order state from the higher-order state and treats the intentional status of the first-order state in the manner that we are familiar with from the intensional context of mental state attributions (Chisholm, 1957). If a subject is in a state *N* which has *M* as its content, then the fact that *N* represents *M* allows for *M* not to exist in exactly the same way as a visual state that represents a plant allows for the plant not to exist. Block’s version of the problem of misrepresentation raises a serious, perhaps fatal, problem for this conception. The alternative non-relational conception holds that in representing *M*, *N* does not represent a state *M* which is separate from *N* in the way just described – *N* casts no beam outwards onto another state – but, rather, it is *N* in which consciousness resides. This conception treats the intentional object in a way that is reminiscent of Brentano’s (2012) theory of immanent intentionality on which the intentional object is contained within the intentional act (see Crane, 2014). However, while this version of the theory may evade Block’s framing of the problem of misrepresentation, as there is no problematic intentional object to which the higher-order state is related, it is subject to a conceptual problem in delineating it as a philosophical theory of consciousness from the first-order theory. This is brought out by considering the original version of the problem of misrepresentation presented by Neander.

The higher-order theory is thus in a dilemma. If it conceives of the theory relationally, then the theory is seriously beset by Block’s version of the problem of misrepresentation. One may hold that the non-relational conception of the theory can avoid this problem, but the non-relational conception runs into conceptual difficulties brought out by Neander’s version of the problem. I will discuss the relational conception in §2 and the non-relational conception in §3.

2. Block and the relational conception

According to the higher-order theory, for the subject to be in a conscious state *M*, *M* has to be the content of a higher-order state *N* through which the subject is (unconsciously) aware of *M*. This results in the subject being consciously in *M* and thus consciously aware of its content *p*. This is a theory of what consciousness is, a theory of what conscious episodes in one’s stream of consciousness, or conscious states in one’s mind, are.⁴ Two

main arguments are presented as motivation for this theory. First, it explains why some of our mental states are conscious and others unconscious.⁵ Second, it accommodates intuitions to the effect that the Transitivity Principle – that conscious states are those that we are aware of being in, or that states of which we are not aware cannot be conscious – is either a conceptual or metaphysical truth, or an element of folk-psychology, and so on.⁶

Unlike the rejection of such arguments for the theory, the problem of misrepresentation does not pose a problem of motivation for the higher-order theory but a direct challenge to its coherence. Block argues that the higher-order theory involves a sufficient and a necessary condition. It is sufficient for a first-order state's being conscious that there be a relevant higher-order state, and necessary for there to be a first-order state that is targeted by the higher-order state. The problem: the sufficient condition locates the source of consciousness in the higher-order state but this is perfectly compatible with a failure of the necessary condition. Block targets the invocation of non-existent intentional objects in response to this problem and argues that applying this in the case of empty higher-order states is implausible. This very response to the problem was already explicitly given by Weisberg (2010), among others, to which Block refers (see also Wilberg, 2010).

“Weisberg says what is conscious is always an intentional object; however sometimes the intentional object does exist and sometimes it does not ... One could put Weisberg's point by saying that there is no such thing as a [higher-order state] with no object because there is always an intentional object that is a conscious state, *though sometimes a nonexistent one!*”. (Block, 2011a, p. 425)

Block conceives of matters on the higher-order theory as follows. In the normal case, where everything functions and represents correctly, the higher-order state N targets the first-order state M , or has M as its intentional object, and thus M is made, or becomes, conscious. In the targetless or empty case, there is no (existent) other state M , but only a state in the sense of a merely intentional object of N . Consider the following situation. I am in a higher-order state that targets a first-order state with the content *that there is a plant on the terrace*. The sufficient condition is satisfied and I therefore have a conscious visual experience of a plant on the terrace. However, as a matter of fact, I am in no such first-order state and the necessary condition that a first-order state be targeted by a higher-order state is not met.

This must be stressed: it is essential that the higher-order state itself is not conscious,⁷ otherwise the theory sets off on a regress (as the higher-order state would have to be targetted by a further third-order state, and so on). Thus, in the present example, as it is my conscious experience of the plant that would be in need of explanation, the higher-order state itself could not

explain this by way of playing the required role in my stream of consciousness. What requires explanation is a conscious experience, and it is difficult to see how an unconscious state could explain this except in a causal sense – but, then, what does it cause?

The response to this problem to which Block objects is that the higher-order state has an intentional object, the first-order state, which although it does not exist is still an intentional object in the way that first-order hallucinations, for example, have intentional objects, and that this non-existent first-order state is the conscious state. Block shows understandable surprise at this position. How could a non-existent state be conscious? Consider a hallucination of a plant on the first-order theory. What needs to be explained is the conscious state, and the first-order theory explains this as a state with a certain content and functional role. As the state merely represents a plant, and representational properties are physical-functional properties, the old worry about the object of the state is dissolved. It was a mistake to wonder where the “plant” is located, how it could be green, etc. In invoking a representational higher-order state, the higher-order theory replicates this move on the higher-order level. While we may be satisfied with this move within the first-order level, there is surely unease with the move made on the higher-order level. While we can explain why any questions about the properties of the plant were misplaced, can the same be said about the first-order state? The property of being a conscious state is not a property that the higher-order state has, but one that results from the higher-order state. In the case under consideration, a first-order state no more exists than does a hallucinated plant. But whereas we can explain the conscious appearance of a hallucinated plant in terms of the properties of an actual, real, existent first-order state, the higher-order theory can offer no such explanation of the consciousness experienced by the subject.⁸

Block expresses the intuitive force of the argument in terms of how states “matter”. How could non-existent consciousness matter, Block asks.

If a theory of consciousness is worth anything, it must be about consciousness in a sense that matters to us in the way that conscious ecstasy or agony matters. If a state of being conscious of agony is supposed to matter equally whether it exists or not, the supposed theory of consciousness is worthless. (Block, 2011a, p. 427)

This intuitive argument has purchase. If I were to hallucinate a plant, the “plant” would matter because of my hallucination: i.e., what would matter would be the state of hallucinating, and its content would determine how it would matter. However, in the case of the empty higher-order state, it is not the higher-order state that matters, but the conscious state. The higher-order state is not itself conscious, and so does not matter in the way that the hallucination of a plant matters. The higher-order state results in a conscious state, a state that matters. The higher-order state only matters

in the sense of being sufficient for a conscious state, not in Block's sense of *being* the state that matters – indeed, it cannot be for the higher-order state is itself unconscious – and mattering in a certain way, depending upon its content. But there is no conscious state to matter. Most vividly, a non-existent pain, for example, could not matter.

The discussion between Block and Rosenthal and Weisberg is mostly expressed in terms of the phenomenology of consciousness, the “what it is like” of conscious experience. Rosenthal and Weisberg seek to rebut Block by arguing that he mistakenly assumes a monadic theory of consciousness. A significant part of back-and-forth concerns whether or not Block assumes that consciousness is to be analyzed as a monadic property. Rosenthal and Weisberg argue that Block assumes that the higher-order state targeting the first-order state (more or less) causes it to have a monadic purely sensational property, a *quale*. Of course, if no such state exists, then Block is right to be suspicious (to say the least) of a non-existent state having this monadic property. A monadic property could not be exemplified by a non-existent state. And even one were desperate enough to cleave to a bizarre form of Meinongianism, a non-existent state to which such a property attached in whatever way one supposes is possible still would not make any mark on the stream of consciousness of a subject who is (not) in the non-existent state.

Rosenthal and Weisberg argue that Block's argument depends on this monadic view of consciousness. Block disagrees, and argues that his point applies to the higher-order theory in any form. In attempting to frame Block as relying on the monadic view, Rosenthal and Weisberg understand Block as interpreting the higher-order theory as follows: there is a first-order state *M* which becomes conscious only when targeted by a higher-order state *N*. The simplest such view (and one which I think that Block does not have in mind) would be that *N* causes *M* to exemplify the monadic property. And, clearly, if *M* does not exist, then this is not possible. While the monadic view is one on which it is clear that the state cannot have the suggested property while not existing, Block presses that his point applies equally to a non-monadic analysis.

It is here that Rosenthal and Weisberg play the intentionality card: if the monadic theory of consciousness is rejected, then if *N* has *M* as its content, *N* can result in the subject being conscious in the way that they would also be were *M* existent without having to posit any more than is required on the first-order level when, for example, a plant is hallucinated. However, as previously noted, when it comes to a conscious state, matters are different than they are for the non-conscious objects of conscious states. The cases diverge because the *explananda* diverge. In one case, we need to explain the mysterious object of a conscious state, and we do so with reference to the properties of the conscious state of which it is an object. In the other case, we have to explain a mysterious conscious state itself. *Here is a conscious state,*

but what are we to say about the plant that seems to be the object of this state? and *What do we say about this conscious state in which there appears to be a plant?* are two different questions that require different answers. Adverting to intentionality as a feature of mental states to successfully answer the former question does not, therefore, allow one to help oneself to the same answer to the latter: the former is about a feature of mental states and how it operates, the latter about the mental states themselves which have this feature which operates in this way.

We can explain a hallucination of a plant, for example, in terms of a misrepresentative first-order state – there is nothing to explain regarding the “plant”, for all the explanation that is required can be given by the hallucination, by the misrepresentative first-order state. Matters are different in the analogous higher-order case. We cannot explain the “internal hallucination”, so to speak, of the first-order state in terms of the higher-order state because that state is unconscious and so cannot provide the explanation of the abilities to act and report provided by the conscious experience which, by hypothesis through the sufficient condition, the subject possesses. And there is, according to the mere intentional object interpretation of the case, also no first-order state that is conscious. So the theory, conceived relationally, lacks the explanatory resources required to deal with the problem.⁹

In arguing that Block relies on the monadic view, Rosenthal and Weisberg correctly diagnose that Block’s argument pushes on a point that is most vividly brought out by the monadic theory of consciousness. But Block is correct that his argument does not rely on the monadic theory of consciousness, but on the general relational theory of which the monadic higher-order theory is a version. If M is a distinct state from N , and M becomes conscious in virtue of its being represented by N , then if N misrepresents M as existing when there is no existent M , then the sufficient condition entails that M is both non-existent and conscious. This is clearly both a troublesome conclusion – troublesome enough to reject the view, in my opinion – and a clear consequence of the view as set out. Yet, if this is so clear, then why do proponents of the higher-order theory push back against Block’s argument? The answer, I believe, is that it is easy to slip between the relational and non-relational conceptions. On the non-relational conception, consciousness resides in the state N , not in a separate state M . There is not a first-order state M that becomes conscious in virtue of being represented by N , but N itself non-causally makes the subject conscious as though they were in a first-order state M . Could recourse to the non-relational conception save the higher-order theory? While it may evade Block’s problem of misrepresentation, consideration of the problem as originally set out by Neander raises some conceptual difficulties for the non-relational conception.

3. Neander and the non-relational conception

On the relational conception, the higher-order state N causes a first-order state M to be conscious, renders or makes M conscious, or bestows consciousness on M . On the non-relational conception, however, N results in the subject being in a conscious state of the type that they would be in were M existent, but without there being an independent M which is acted upon by N . On this view, as I understand it, the higher-order state makes the subject conscious of the content of the ostensible first-order state that is its content without the (constitutive) involvement of a separate first-order state. That is, N is transparent, in the common sense of Moore (1903) and Harman (1990), to the content of M . Just as first-order theories hold that conscious first-order states are transparent to their content, so on the non-relational conception the higher-order state is transparent to the content of the first-order state that it represents.

Note the difference from the relational conception of the theory, typified most clearly by the monadic view. The higher-order state N has as its content the state M which has a content p , and in virtue of being in N , the subject is conscious that p – just as they would be were they in M . *Not* that the subject is *in* state M in virtue of being in N and thus through M conscious that p . The subject is unconsciously in misrepresentative state N , which makes them erroneously (unconsciously) aware of a state M and thus it is to them *as though they were in* conscious state M – i.e., that they are consciously aware that p – even though they are not as they are only in N .

This seems to me to be essentially the move made by Brown (2015), Lau and Brown (2019), and Berger (2014), in addition to Rosenthal and Weisberg. Here are Rosenthal, Weisberg, and Brown on this point.

What counts for somebody's being in a conscious state is just the occurrence in one's stream of consciousness of the relevant subjective appearance, the appearance of being in the state in question. A theory of consciousness explains conscious subjectivity, not the underlying mental reality that's responsible for these appearances. (Rosenthal, 2011, p. 432)

According to the [higher-order] approach, when there's something it's like for one, it's due to the subject's being aware of her states by way of [higher-order] representation. In the misrepresentation case, she is aware of herself as being in a state that happens to be absent. So what it's like for her is just as she (erroneously) represents things as being. If she seems to herself to be in pain, that's what it's like for her. (Weisberg, 2011, p. 440)

On this alternative way of thinking about the higher order theory there is no explanatory role for a relation between first and higher-order states. Rather the explanatory power lies in the nature of the higher-order representation in question . . .

In the case of consciously seeing red, for instance, the higher order representation will be something to the effect of ‘I am having this visual representation of red’ (Brown, 2015:, 2015).

Thus one might hold that Block’s objection misses the point. The higher-order state does not make a first-order state conscious, but makes the subject conscious as they would be were they in the first-order state – both when it exists and when it does not exist. However, this version of the view is subject to a conceptual problem in delineating it from the first-order theory as a philosophical theory of consciousness. We can draw this out by considering Neander’s original version of the problem of misrepresentation.¹⁰ She argues that the higher-order theory involves a “division of phenomenal labor” between the first-order state that supplies how things appear and the higher-order state that supplies the consciousness of this appearance. This, essentially, results in the first-order state being no more than a spinning wheel in an explanation of consciousness.

The proposed division of phenomenal labor seems to imply representational action at a distance. Sensory processing produces representations that are supposed to be consciously experienced only if our self-monitoring system monitors them and in doing so makes representations of them . . . So what makes the difference between a conscious and an unconscious sensation is, we are being told, in effect what goes on in ‘down-stream’ processing. But, given that down-stream processing is needed for (and presumably before) the sensory representation is consciously experienced, it is hard to see how the original sensory representation can keep any independent hold, as it were, on *how* it is consciously experienced. If *whether* it is experienced depends on down-stream processing, *how* it is experienced should surely depend on down-stream processing, since by then it is only present by proxy (i.e., by representation). So if the second-order representation misrepresents it at the point where the information about it has been integrated enough for it to be consciously experienced, it is quite unclear how it can still, from that computational distance, as it were, control how it is experienced (Neander, 1998, pp. 424–425).¹¹

If the higher-order state determines whether or not the subject is conscious, it will also determine in what way they are conscious – i.e., what content they are conscious of. The misrepresentation problem brings this clearly out: if being in the higher-order state *N* is a sufficient condition for undergoing a conscious experience, then whether or not there is a separate state *M* which preceded it in the causal chain which resulted in the subject being in *N* is irrelevant to how things appear to the subject in *N*. In explaining consciousness, there is no explanatory, non-causal, need for a distinct, prior first-order state in one’s explanation of consciousness. As *M* comes before *N* in the causal chain and is only present from the perspective of *N* by being represented in *N*, it is the representational content of *N* that controls how things consciously appear for the subject

and not *M*. *M* plays a part in the causal explanation of the subject's conscious state but not in the explanation of what this conscious state is: that explanation is exhausted by *N*.

Just as with Block's objection, one might argue that Neander's objection applies only to the relational conception.¹² The non-relational conception concedes at the outset that the first-order state merely supplies the higher-order conscious state with its content and causes it in the correct way. However, consideration of the spinning wheel objection brings out a conceptual problem for the non-relational conception. Consider the non-relational theory as presented here by Brown.

If one were consciously experiencing a nice vivid red then the relevant higher-order representation in question will have something like '*I am, myself, having this visual representation of red*' as its intentional content. If one were consciously seeing green instead then the content would be something like, '*I am, myself having this visual representation of green*' . . . the difference between a phenomenally conscious experience of green and one of red is in the content of the higher-order representation not in the first-order states that one instantiates (Brown, 2015; 2015).

When I consciously see red, for example, this is because I am instantiating a suitable higher-order representation but I am not aware of myself as being in that higher-order state. In order for that to happen I would have to instantiate a further state representing me as being in the previous state, which is not what happens in ordinary perceiving (Brown, 2015; 2015).

Neander's point against the higher-order theory is that any talk of a first-order state is idle in an explanation of consciousness – it is the higher-order state that accounts for consciousness. This is clearly true on the non-relational conception. The higher-order state has a content, e.g., *that I am experiencing a vivid red*, and the content of this state, the vivid red, is what the subject is conscious of. There is no role for a separate first-order state with this content, i.e., the vivid red, in the explanation of consciousness – there is one state, the higher-order state, and this state is conscious in virtue of its being the type of state that it is in having the content and functional role that it has.

But consider how the theory now under consideration is to be distinguished from the first-order theory as a theory of what consciousness is, or of what conscious states are, when one also takes into account the fact that on the higher-order theory as presented one is not conscious of, or aware of being in, *N*, nor of the *state M*, but only of the *content* of *M*. The subject does not have a conscious experience of the *state M*, only of its content: it is not for them consciously that *I am experiencing a vivid red*, but rather only that *there is a vivid red*.

Consider the following simplified explanation of the first-order theory. Conscious experiences are states with a certain first-order content and functional role. These states are transparent to this first-order content. These states have causal antecedents that represent (parts of) the same things as the conscious states do and which (together) cause the conscious states to have their content when the first-order states are formed. The conscious states also represent or carry information about other things, presumably including these states, only the subject is not aware of these contents. They are sub-personal representations. So, we have two categories of representation, i.e., things represented, with respect to conscious states: the contents that the state brings before the subject's mind in conscious experience, and the contents that are not brought before the subject's mind, including representations of the proximate causal antecedents.¹³

So, a conscious state M will have a causal role, R , which specifies how it mediates perceptual input and cognitive and behavioral output, a first-order content p which is transparently brought before the subject's mind, and a further range of contents q_1 to q_n which it represents but which are not brought before the subject's mind and which therefore remain at the sub-personal level. Among these contents will be the proximate causal antecedents. But compare this now to the higher-order theory as now under consideration. The conscious state N has a certain functional role, R , which specifies how it mediates perceptual input and cognitive and behavioral output. The conscious state has a first-order content, p , that it transparently brings before the subject's mind, and a range of contents q_1 to q_n which it represents but which are not brought before the subject's mind and which therefore remain at the sub-personal level. And here is the point: is the state M , not the content p of M but M itself, not merely one of the contents q_1 to q_n ? M is represented by N but is not brought before the subject's mind. The subject is not conscious of M , or (directly) conscious that they are in M , only conscious of its content p . Further, the causal roles will be the same with respect to how the conscious state mediates perceptual input and cognitive and behavioral output.¹⁴ In a discussion of the higher-order theory, Gottlieb (2016) raises a question in the vicinity.

Suppose we did parse [the transitivity principle] as the claim that, in virtue of being in a state of higher-order awareness, we are (merely) aware of what our first-order mental states represent . . . Then, we would be forced to say that our states of higher-order awareness don't have higher-order content. In other words, the proposal would collapse into absurdity. Higher-order states might also represent what our (first-order) conscious states represent (e.g., redness, trees), but it cannot only represent these things. So while we can describe [the transitivity principle] as the claim that conscious mental states are mental states we are aware of as having certain contents, this cannot be shorthand for the claim that we are just aware of what contents our

conscious states have. A First-Order theorist like Tye or Dretske could accept that reading of [the transitivity principle], as there is nothing here that requires a state of higher-order awareness. (Gottlieb, 2016, p. 358)

This is a conceptual problem for the proponent of the non-relational conception. Not only could the first-order theorist accept this understanding of what conscious states are, but this *is* their understanding of what conscious states are. If we understand by the higher-order theory that the higher-order state results in the content of the (ostensible) first-order state being consciously brought before the subject's mind because it is, so to speak, a nested content of the higher-order state, then there is nothing of great substance that the first-order theorists will disagree with *qua* a philosophical, or conceptual, account of what consciousness is. A conscious state is one which has a certain functional role and a first-order content of which the subject is conscious in virtue of the state's functional role. We understand content here as the content that the subject can (in normal cases) immediately report and act upon, not merely as anything wider that is represented by the state but not consciously represented to the subject – for example, the content of a proximate state. The latter understanding of content is not what the higher-order theory understands content to be in this context, for this does not isolate the first-order state as a content of note; and the former understanding of content excludes the first-order state but includes its content and so is what the first-order theorist proposes. Here the spinning wheel metaphor is again raised. What conceptual work is *M* doing on this theory? It seems that *M* could be jettisoned from an account of what consciousness is, or what conscious states are. Perhaps there may be some need for *M* in the full representational account of the implementation of functional role *R* with respect to content *p* in *N*. But it is not clear that this will be objectionable to a proponent of a first-order theory of consciousness who will also accept that there is more to the substrate which implements the conscious state than that which is accounted for by the access-conscious properties of the state. In order to implement the functional role *R* with respect to *p* in the conscious state, the vehicle or realizer may have a complicated wider functional role. However, the theory as presented is not one about which we need say that the subject is in a conscious state *M* because they are in a state *N* through which they are unconsciously aware of *M*. Rather, they seem just to be in one state in which a neural vehicle realizes the functional role of making them access-conscious that *p*. That vehicle may have further sub-personal representational properties, but that is not something to which the first-order theorist disagrees, nor is it a cognitive requirement of consciousness – i.e., a cognitive condition on a state being conscious. Indeed, as the structure of the conscious state is

that there is just one state whose functional role renders the subject conscious of a first-order content of that state, it is not immediately clear why the vehicle, or realizer, which implements the conscious state should be described in terms of two states, *M* and *N*. There are not really two different cognitive states, i.e., a first-order state with first-order content and a second-order state with second-order content, with a causal relation between them in virtue of which one of the states is conscious. There is just one state in which, in virtue of its functional role which is realized by a vehicle with wider sub-personal representational properties, the subject is conscious of a first-order content of that state. Let me consider two potential responses to this argument.

In the face of this problem, one could argue that in being in the higher-order state one is conscious not just that *p*, the content of the first-order state, but conscious of the first-order state itself along with its content – i.e., that the subject is conscious that *p* and conscious that they are in the state *M* of being conscious that *p*. But, as Block notes, this at least takes the view very close to the reflexive theory, not the higher-order theory. The distinction between the higher-order and reflexive theories is not quite sharp, but if a theory holds that the subject is conscious of both the first-order content and the state which has the first-order content, and that these are the same state, then the theory must surely be classified as a version of the reflexive and not the higher-order theory.

The resulting theory looks like the reflexive theory because on the non-relational conception the first-order state is conceived to be (in some sense) a part of the higher-order state. On this view, the higher-order state does not represent a distinct state, i.e., one which is not a part of itself, which stands in causal relations to it, and which may or may not be obtained. Rather it contains the first-order state as (in some sense) a part, and so the state represents a part of itself. Unless a condition on the reflexive theory as opposed to the higher-order theory is that on the former conscious states must represent themselves in their entirety, a condition that seems *ad hoc*, this seems to me to render a theory on which a conscious state is conscious in virtue of representing a part of itself a version of the reflexive theory.

It does not follow from the foregoing arguments that retreat to a version of the reflexive theory is fatal for the higher-order theory – if, that is, the resulting version of the reflexive theory is correct.¹⁵ However, if the foregoing arguments are correct, it is important to note two things. First, the higher-order theory has been boxed into this corner. Block's argument is fatal for the relational conception, and Neander's argument brings out a conceptual problem in holding that conscious states are transparent to first-order content. Thus, the proponent of the higher-order theory must hold a version of the reflexive theory that subscribes to the controversial claim – note, one that is certainly not typically presented as part of the

higher-order view but is rather typically presented as a constitutive part of a reflexive theory – that conscious states are states in which the subject is conscious not just of the first-order content but also of the state itself. This leads to the second point. We have traveled far from the intuitive conception of the higher-order theory, as typified by the relational conception. Instead of a first-order state being conscious because it is the object of a state of higher-order awareness, a claim no part of which is that the subject is conscious of the state itself as opposed to its content or that the first-order state obtains when the higher-order state obtains, we now have a view on which there is one state that represents a part of itself, a part that necessarily obtains if the state itself obtains, and in doing so makes the subject conscious of both first and second-order content. This seems to render the view a version of the reflexive theory. The reflexive theory, with these further commitments, has a number of problems of its own.¹⁶

Perhaps, then, to avoid recourse to the reflexive theory, the proponent of the higher-order theory may point to empirical results in support of their theory. It is currently a matter of debate in the literature as to where the neural correlates of visual conscious experience are located (e.g., Block, 2007; Kozuch, 2014; Lau & Brown, 2019; Lau et al., 2019; Michel & Morales, 2019). First and higher-order theorists propose different locations for the neural correlates of visual consciousness. Let me take Lau and Brown (2019) as an example. Lau & Brown defend a higher-order theory on which consciousness results from (higher-order) representations in the prefrontal cortex of early visual (first-order) representations (the preferred location of the neural correlates of conscious vision for the first-order theorist). They discuss empirical cases where there seems to be conscious experience without the early visual representations. One of the cases Lau & Brown present, Rare Charles Bonnet Syndrome, involves subjects who report visual sensations despite lacking the states assumed by the first-order theorist to be neural correlates. There is some controversy over the interpretation of the empirical facts in these cases and the suitability of experiments that require reports of conscious experience as it can be argued that the correlates in these cases are correlates of reports not consciousness. Nevertheless, these cases are raised against Block's neural vehicle view of consciousness, for example, as it is the first-order neural vehicle that is lacking in these cases. The point would extend to first-order representational theories with the same commitments regarding the location of neural correlates. But even if we grant this interpretation of the empirical facts, and we accept the experiments involved, there is still the conceptual question to be asked about how distinct the higher-order theory is as a theory of what consciousness is, as opposed to where consciousness is located.

As Lau, Brown, & LeDoux note, '[the higher-order theory] is not about [the pre-frontal cortex] *per se*. It is a theory about fundamental cognitive requirements of consciousness' (2019: 761). We can separate two questions. First, what are the neural correlates of consciousness. Second, assuming that we reject a type-identity theory of consciousness on which conscious states are simply identical to their neural correlates with no more to be said, what are the conscious states of which these are the neural correlates – in Lau, Brown & Le Doux's terms, what are the 'fundamental cognitive requirements of consciousness'? Here, the higher-order theory proposes the higher-order answer, namely that conscious states are those which are represented by the appropriate higher-order state. On the relational conception of the theory, this cognitive requirement is clear. However, the relational conception is, it seems to me, sunk by Block's argument. This leaves the non-relational conception. And the spinning-wheel problem seems to call into question whether the non-relational conception of the (non-reflexive) higher-order theory genuinely offers a higher-order account of these fundamental cognitive requirements, for the cognitive requirements suggested – the functional role in mediating input and cognitive and behavioral output and the content to which this applies – are shared by the first-order theory. Disagreements between proponents of the first-order theory and the (non-reflexive) non-relational theorists concerning the location and structure of the neural correlates will be disagreements over just that, not substantive disagreements over the metaphysics, or fundamental cognitive requirements, of conscious states. The relational conception of the theory does provide a substantive alternative to the first-order theory in the respect. On this view, conscious states are part of a cognitive nexus in which they are integrated with higher-order states that feed back to render them conscious in virtue of this nexus reflecting the underlying truth of the Transitivity Principle. On the non-reflexive, non-relational view, however, the disagreement will concern only the sub-personal structure of the neural correlates of conscious states. And this is not a disagreement over the cognitive requirements, or metaphysics, of consciousness itself. Here, the two theories seem to be in agreement: conscious states are those in which the subject is conscious of a first-order content in virtue of the functional role of the state. They may disagree over what is required to implement this functional role (though without a clear adherence to the Transitivity Principle as found in the relational version, it is perhaps not immediately clear why), but they will not disagree over what this role, i.e., over what consciousness, is.

4. Conclusion

Theories of consciousness are theories of phenomenal consciousness, even those, like the functional-representational theory, which reduce or eliminate this notion in functional-representational terms. The higher-order theory is a species of this view, and so offers an account of the fundamental cognitive requirements of consciousness in these terms. There are three options for the higher-order theory. The first is the relational conception. On this view, a first-order state becomes conscious, or is caused to be conscious, once the content of that state is integrated into higher cognitive functions. Here we have a genuinely higher-order account of the cognitive requirements of consciousness. However, this view faces the significant problem raised by Block. Thus, the higher-order theorist must retreat to the non-relational conception. And here they face a choice. Is the subject conscious just of first-order content or also of the first-order state when they are in a conscious first-order state? If the latter, then the theory on offer is a version of the reflexive theory, one with different commitments than the higher-order theory as typically presented. The viability of this theory is, of course, a further question. If the former, then how does the account of the cognitive requirements of consciousness differ from than of the first-order theory? For the account is that first-order conscious states are those with a certain functional role in which a first-order content is access-conscious. If there are cases such as those set out by Lau & Brown, then a first-order theory that locates conscious states in the early visual system will be in trouble (as, to note, would a relation conception of the higher-order theory that shares this location for conscious first-order states). This might therefore rule out theories such as Block's biological first-order theory and theories that agree about the location of the neural correlates. However, if offered in a defense of a non-relational higher-order theory as a philosophical theory of consciousness, of what conscious states are, then this non-relational higher-order theory will not in fact suggest anything that a proponent of a first-order functional representational theory will disagree with.

The problem of misrepresentation thus presents a significant dilemma for the higher-order theory of consciousness. There are two conceptions of the higher-order theory and two versions of the problem of misrepresentation. Block's version regarding the implausibility of non-existent conscious states raises a serious difficulty for the relational conception of the higher-order theory of consciousness. But consideration of the spinning-wheel objection from Neander draws out that it is questionable whether the non-relational conception of the theory presents a genuinely higher-order philosophical theory of consciousness. Unless, that is, the non-relational theory is understood to be a version of the reflexive theory.

Notes

1. For example, Dretske (1996), Harman (1990), and Tye (1995). Some theories are “first-order” without being representational in that they hold that it is the physical substrate of the first-order representations which determines consciousness, not the functional role of the state with these representations (e.g., Block, 2009). Usually, though, problems for the higher-order theory are taken as arguments in favor of another version of the representational theory as anti-representationalists tend to reject the general representational framework for independent reasons.
2. Some versions of the theory are actualist in that the first-order state has to actually be the content of a higher-order state (e.g., Gennaro, 2012; Lycan, 1996; Rosenthal, 1986). Others are dispositionalist, in that the first-order state has to only be disposed to be the content of a higher-order state (e.g., Carruthers, 2000). The former is the more popular view, and the arguments in this paper apply to this version. A further distinction is between those theories which hold the higher-order state to be cognitive (e.g., Rosenthal) and those which hold it to be perception-like (e.g., Lycan). This distinction does not matter for this paper. I will sometimes talk of the higher-order state as a state of “awareness” of the first-order state, and this is intended to be neutral.
3. For example, Kriegel (2009). Although the reflexive theory is sometimes taken to be a version of the higher-order theory, I would classify it as a different theory due to its separation of two independent states which have the two different contents. See Peebles (2018) and §3 for some discussion of this issue.
4. Block (2011a) calls this the “ambitious” version of the view which is to be contrasted with the “modest” version of the view which would only tell us about higher-order consciousness, not consciousness *per se*.
5. See, for example, Lau and Rosenthal (2011), Lau and Brown (2019).
6. See Rosenthal (1986) for a famous statement of such a view and Lycan (2001) for a succinct statement. Despite such influential remarks as “I cannot myself hear a natural sense of the phrase ‘conscious state’ other than meaning ‘state that one is conscious of being in’” (Lycan, 1996, p. 25), I, like many philosophers, do not agree with this. At least I do not agree with this if unpacking “state that one is conscious of being in” results in any metaphysical, conceptual, or folk-psychological commitments regarding the nature of consciousness. However, this argument is not the subject of this paper.
7. In the normal case of mere first-order consciousness, that is. Nothing precludes the higher-order state from becoming conscious through introspection, for example.
8. The problem cannot be avoided. The empty higher-order state is an odd case, but it is one that the theory allows for. Indeed, it is the fact that the representational theory allows for non-existent intentional objects that makes it so attractive in the first place. Otherwise we would still be worrying about where the hallucinated plant is and what properties it has (or worrying about something similar like qualia). It is therefore a nonstarter to deny the possibility of empty or misaligned higher-order states. The representational theory of mind supplies the innocent metaphysics of merely intentional objects and this necessarily involves the possibility of misrepresentation which brings with it this particular problem.
9. I believe that this is a good way to put the point. Berger (2014), for example, attacks Wilberg’s (2010) version of this response to the problem, which is to accept the sufficient condition and hold that the conscious state does not exist, but defend this by holding that just as we can have, using my example, the appearance of a plant without there being a plant (in hallucination), so we can have the appearance of

a conscious state without there being a conscious state. Berger rightly notes that ‘our ordinary conception of conscious arguably regards this kind of subjective impression as sufficient for consciousness’ (2014: 833) and goes on to press the point. The appearance/reality distinction is *prima facie* questionable in the case of consciousness (unlike in cases of the appearance of other things such as plants, in this case we are dealing with appearance of appearances). However, the simple point is easy to grasp without embroiling us in a metaphysical discussion about consciousness and appearances. A subject in a targetless state can, by hypothesis, describe the resulting experience. They can describe its features, compare it to other states, and so on. It seems difficult to deny that the subject is therefore in a state which is conscious. If they are not, then what kind of state is this: what are its properties; why, when these properties “feel” just like conscious properties, are they not just conscious properties; and so on? Indeed, this is just the consequence of the sufficient condition. The proponent of the view therefore must answer the question: with what is the state responsible for these abilities to be identified? And to this question they do not have a satisfactory answer. This particular way of raising the problem also avoids extended discussion of whether the states that result from targetless higher-order states would possess or lack the usual functional role and therefore behavioral effects. Lycan (1996, p. 20) and Rosenthal (2011, p. 436) argue, for example, that such pain experiences would lack the usual behavioral effects. The question of whether these can be said to be genuine conscious experiences is then raised. One feature to note of Lycan and Rosenthal’s claim is that it separates the feel of conscious pain experiences from the functional role of pain experiences. And this is not something that most proponents of the functional-representational theory are likely to be sympathetic to. Further, and to return to the main point, the original question still stands in need of a satisfactory answer regardless of whether the state in question lacks the usual behavioral effects. With what is it to be identified?

10. Neander considers the general case where the two contents misalign. Block puts the argument slightly differently, focusing on the empty case. The empty or targetless case is more vivid, but the same problems apply to all cases of misrepresentation.
11. Neander’s subject is Lycan’s (1996) representative version of the theory. She extends her argument to cover what she calls “broad” versions of the first-order theory, such as Dretske (1996). Most first-order representationalists would consider themselves to subscribe to some version of this theory and so the reader might therefore expect some comment on this part of her argument, but to do so would take the course of this paper too far afield.
12. I do not want to enter into a discussion of whether or not it is fatal for the relational conception (the problems arising from Block’s argument are enough to consider the relational conception to be in difficulty) but it is another significant problem for the relational conception.
13. I assume a generic and nonspecific account of representation in terms of “carrying information” because the argument is intended to be neutral on a metaphysics of representation. It should not be affected by the adoption of any more specific account of representation.
14. Rosenthal argues that “Though the first-order sensation plays no noncausal role in determining what it’s like for one, the sensation is not [on his account] mentally inert” (2018: 54). He argues that the first-order state plays a role in unconscious priming, Stroop effects, and reaction times, among other testable psychological phenomena: “the first-order state figures in psychological functioning other than consciousness, and the higher-order state constitutes subjective awareness” (2018: 54). However,

note that Rosenthal's suggestions for the causal role of the first-order state are all unconscious functions. There is nothing for the first-order theorist to object to if the claim is that unconscious states can have an effect on conscious behavior. A stronger claim, one which places conscious functions on the first-order state pushes the view toward the relational conception, and thus into Block's problem.

15. I would like to thank an anonymous reviewer for pressing this point.
16. See Peebles (2018) for further discussion of the reservations about this sort of claim and the reflexive theory in general.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Berger, J. (2014). Consciousness is not a property of states: A reply to Wilberg. *Philosophical Psychology*, 27(6), 829–842.
- Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences*, 18(2), 227–247.
- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30(5–6), 499–548.
- Block, N. (2009). Comparing the major theories of consciousness. In M. Gazzaniga (Ed.), *The cognitive neurosciences IV* (pp. 1111–1122). MIT Press.
- Block, N. (2011a). The higher order approach to consciousness is defunct. *Analysis*, 71(3), 419–431.
- Block, N. (2011b). Response to rosenthal and weisberg. *Analysis*, 71(3), 443–448.
- Brentano, F. (2012). *Psychology from an empirical standpoint*. Routledge.
- Brown, R. (2015). The HOROR theory of phenomenal consciousness. *Philosophical Studies*, 172(7), 1783–1794.
- Byrne, A. (1997). Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies*, 86(2), 103–129.
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Oxford University Press.
- Chisholm, R. (1957). *Perceiving: A philosophical study*. Cornell University Press: Ithica.
- Crane, T. (2014). Brentano's concept of intentional existence. In *Aspects of psychologism* (pp. 25–39). Harvard University Press.
- Dretske, F. (1996). *Naturalizing the mind*. MIT Press.
- Gennaro, R. (2012). *The consciousness paradox: Consciousness, concepts, and higher-order thoughts*. MIT Press.
- Gottlieb, J. (2016). Transitivity and transparency. *Analytic Philosophy*, 57(4), 353–379.
- Harman, G. (1990). The intrinsic quality of experience. *Philosophical Perspectives*, 4, 31–52.
- Kozuch. (2014). Prefrontal lesion evidence against higher-order theories of consciousness. *Philosophical Studies*, 167(3), 721–746.
- Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford University Press.

- Lau, H., & Brown, R. (2019). The emperor's new phenomenology? The empirical case for conscious experience without first-order representations. In A. Pautz & D. Stoljar (Eds.), *Blockheads! Essays on Ned Block's philosophy of mind and consciousness* (pp. 171–198). MIT Press.
- Lau, H., Brown, R., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences*, 23(9), 754–768.
- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15(8), 365–373.
- Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford University Press.
- Lycan, W. G. (1996). *Consciousness and experience*. MIT Press.
- Lycan, W. G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, 61(1), 3–4.
- Michel, M., & Morales, J. (2019). Minority reports: Consciousness and the prefrontal cortex. *Mind & Language*, 1–21.
- Moore, G. E. (1903). The refutation of idealism. *Mind*, 12(48), 433–453.
- Neander, K. (1998). The division of phenomenal labor: A problem for representationalist theories of consciousness. *Philosophical Perspectives*, 12, 411–434.
- Peebles, G. (2018). Reflexive theories of consciousness and unconscious perception. *Philosophical Psychology*, 31(1), 25–43.
- Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies*, 49(3), 329–359.
- Rosenthal, D. (2011). Exaggerated reports: Reply to block. *Analysis*, 71(3), 431–437.
- Tye, M. (1995). *Ten Problems of Consciousness*. MIT Press.
- Weisberg, J. (2010). Misrepresenting consciousness. *Philosophical Studies*, 154(3), 409–433.
- Weisberg, J. (2011). Abusing the notion of what-it's-like-ness: A response to block. *Analysis*, 71(3), 438–443.
- Wilberg, J. (2010). Consciousness and false HOTs. *Philosophical Psychology*, 23(5), 617–638.