

Higher-Order Theories of Consciousness

Josh Weisberg

The Oxford Handbook of the Philosophy of Consciousness

Edited by Uriah Kriegel

Print Publication Date: Jul 2020 Subject: Philosophy, Philosophy of Mind Online Publication Date: Jul 2020

DOI: 10.1093/oxfordhb/9780198749677.013.20

Abstract and Keywords

Higher-order (HO) theories of consciousness hold that a mental state is conscious when it is appropriately represented by a ‘higher-order’ state, a state about another mental state. The higher-order perception (HOP) theory holds that HO representation is best modeled on perceptual processes, while the higher-order thought (HOT) theory holds that it is best modeled on thought. In addition, some HO theories hold that to be conscious, a state must be actively represented by an HO state, while others maintain that the mere disposition to be represented by an HO state is enough. The HO theory, if successful, offers a reductive explanation of mental state consciousness in terms of nonconscious HO representation. This chapter first spells out the general motivation for the HO view and the differences between HOP and HOT before considering key objections to the approach, as well as possible empirical support. Finally it looks at how the view addresses the explanatory gap and the hard problem of consciousness.

Keywords: Higher-order theory, higher-order perception, higher-order thought, consciousness, theories of consciousness, sensory qualities, phenomenal consciousness, hard problem of consciousness, explanatory gap

Higher-order (HO) theories of consciousness hold that a mental state is conscious when it is appropriately represented by a ‘higher-order’ state, a state about another mental state. The higher-order perception (HOP) theory holds that HO representation is best modeled on perceptual processes, while the higher-order thought (HOT) theory holds that it is best modeled on thought. In addition, some HO theories hold that to be conscious, a state must be actively represented by an HO state, while others maintain that the mere disposition to be represented by an HO state is enough. The HO theory, if successful, offers a reductive explanation of mental state consciousness in terms of nonconscious HO representation.¹ In this chapter, I will first spell out the general motivation for the HO view and the differences between HOP and HOT. Then I will consider key objections to the approach, as well as possible empirical support. I will close by looking at how the view addresses the explanatory gap and the hard problem of consciousness.

20.1 The Transitivity Principle

20.1.1 The Transitivity Principle Explained

HO theories are motivated by what David Rosenthal (2000) terms the ‘transitivity principle’, the idea that *conscious states are states we’re conscious of*. On the face of it, the (p. 439) principle looks circular. However, the term ‘consciousness’ has both a transitive and intransitive use. In the transitive sense, we are conscious *of* something; in the intransitive sense, a mental state is either conscious or nonconscious. The transitivity principle characterizes the intransitive consciousness of a mental state in terms of a subject’s transitive consciousness *of* that mental state.² There are plausible explanations of transitive consciousness in terms of representation, explanations independent of intransitive mental state consciousness.³ So it is open to the HO theorist to explain intransitive mental state consciousness in terms of representational transitive consciousness.

But what reason is there to accept the transitivity principle? First, the principle is folk-psychologically plausible, especially in its contrapositive form: we would intuitively deny that a state is conscious if we are in no way aware of it.⁴ This seems to point to a folk-theoretic way of distinguishing conscious and nonconscious states: implicit in our folk psychology is the idea that conscious states are states we are aware of. If the initial data for theorizing about the mind is given by a ‘rough and ready’ analysis of our everyday folk-psychological concepts,⁵ then the transitivity principle plausibly captures at least one everyday sense of ‘consciousness’. Second, we can appeal to phenomenology. First-person reflection may reveal that consciousness possesses the sort of inner- or self-awareness characterized by the transitivity principle. When I am aware of my backyard, for example, there also seems to be a kind of awareness of myself as seeing the back yard, or an awareness that I am seeing it. This awareness may be more or less focused and it may even fade into the background when I am fully absorbed in a task. Still, it is claimed, this ‘for-me-ness’ is always present in experience to some degree. Different conscious states may involve the rich colors of my backyard, or the pain in my upper back, or even my troubled reflections on the hard problem, but they will share the kind of reflexive awareness characterized by the transitivity principle.⁶ Third, one might appeal to the theoretical utility of embracing the transitivity principle. It may provide a useful way of empirically individuating conscious states—in experimental psychology, for example, taking conscious states to be states we are aware of being in may facilitate experimental design or avoid conceptual confusion. Or it may provide the best overall theoretical ‘fit’ for conscious states, given our commonsense, empirical, and metaphysical commitments. That is, the task of finding a global equilibrium between these various concerns may best be served by embracing the transitivity principle. Fourth, the transitivity principle may best connect a theory of consciousness to certain historical antecedents, like the views of Aristotle, Locke, or Sartre.⁷ And finally, it may be that the transitivity principle provides the best opening wedge to reductively explaining (p. 440) consciousness. Remaining explanatory worries might then be tackled within the framework developed to capture the transitivity principle. Taken together, these reasons provide support for the claim that the transitivity principle appropriately fixes the data a theory of consciousness must explain.

Two refinements of the transitivity principle are needed. In its basic form, the principle states that we are conscious of our conscious states. But we can be conscious of our mental states in ways that fail to make those states conscious. If I am in a state of anger and I am informed of this fact by someone I trust, I may come to be conscious of my anger. But that may not make the anger itself conscious. I may still be in denial or otherwise unable to directly access the state. My awareness of the state must, therefore, occur in a seemingly direct way. I cannot be aware of any mediating inference or observation leading to the awareness—it must seem from the first-person perspective to be spontaneous and unmediated. This is captured in the transitivity principle by adding ‘appropriately’: a mental state is conscious when one is *appropriately* aware of it. The right way to cash out this sort of awareness is a matter for the theory to determine, but it is clear that not just any awareness of one’s state will do.

A second refinement concerns whose state one is aware of. It seems that even if I become aware of one of your states in a seemingly direct way, that will not make your state conscious. I must be appropriately aware of *myself* as being in the state. So some sort of self-reference needs to be captured in the transitivity principle. This delivers the full transitivity principle:

TP: A mental state is conscious when one is appropriately conscious of oneself as being in that state.

This characterization of consciousness is meant to be pretheoretic, offered prior to the theoretical positing of any mechanism that might explain how this process is instantiated in us. HO theorists take the transitivity principle as fixing the data a theory of mental state consciousness must explain and develop their theories accordingly.

20.1.2 Challenges to the Transitivity Principle

The use of the transitivity principle in fixing the data can be challenged in a number of ways. Charles Siewert (1998: sect 6.3) argues that while there is a sense in which we use ‘conscious of’ language to refer to an awareness of our conscious states, more often we use the phrase to report awareness of objects in the world.⁸ This suggests that we should not rush to take ‘conscious of’ talk as implying anything about state- or self-awareness. Instead, it normally marks off the worldly content we are aware of. But the HO theorist can accept this claim without giving up the idea that the transitivity principle properly (p. 441) fixes the data. The principle is not justified on the basis of frequency of use; rather, it is justified as a (likely implicit) folk-psychological distinction between conscious and unconscious states. The intuitive difference between a conscious and nonconscious visual representation of a ball is that we are suitably aware of ourselves as seeing the ball in the conscious case. And this justifies the transitivity principle, even if there are many contexts where we use ‘conscious of’ to pick out things other than our own states.

Fred Dretske (1993) directly challenges the transitivity principle by presenting a case allegedly showing that we can be in a conscious state we are not conscious of. He calls this the ‘spot’ case. Consider an array containing thirteen similar spots in a random pattern. We might look at the array and have a conscious experience of it. Later, we might be presented with an array with only twelve of the previously seen spots and consciously experience that as well. However, we often will not be aware of any difference between two arrays, even though one spot is missing. Dretske contends that the experience of the thirteenth spot must have been conscious, because we were looking at in good light, attending to the array, etc. But we are not conscious of the difference between our two experiences. It follows that we are conscious of the thirteenth spot even though we are not conscious of our experience of it. Thus, the transitivity principle is wrong. However, the defender of the transitivity principle can hold that we can be conscious of our states without being conscious of them in every respect. Importantly, we might be conscious of our states but not conscious of the property that makes the difference between two experiences. We might even be conscious of the relevant spot, but not under the description of ‘difference maker’. So the spot case does not contradict the transitivity principle.⁹

More pressing is the charge that the transitivity principle, while reasonably picking out one kind of consciousness, is not about the one most relevant to the study of consciousness: consciousness characterized in terms of ‘what it’s like for the subject’.¹⁰ In response, proponents of the transitivity principle point to Thomas Nagel’s initial ‘what it’s like’ characterization of consciousness (Nagel 1974). Nagel writes that ‘an organism has conscious mental states if and only if there is something that it is like to *be* that organism—something it is like *for* the organism’ (Nagel 1974: 436, emphasis in original). Nagel himself stresses that there must be something it is like *for* the organism. Defenders of the transitivity principle argue that the best explanation of there being something it is like *for* the organism is that the organism is aware of its states. If we are in no way aware of being in a state, there is noth-

ing it is like *for us* to be in that state. So, the transitivity principle does not leave out what it is like; indeed, it makes clear just why there is something it is like for us when we are in conscious states.

But recently, Daniel Stoljar has challenged this ‘reflexive’ reading of Nagel’s ‘what it’s like’ characterization (Stoljar 2016). He argues for what he terms the ‘affective approach’ (p. 442) to interpreting ‘what it’s like’ talk in general and Nagel’s phrase in particular. Stoljar’s analysis of Nagel’s phrase is as follows:

For any subject S and any psychological state X of S, X is a phenomenally conscious state if and only if X is constitutively such that there is some way that S feels in virtue of S’s being in X (Stoljar 2016: 1190).

Stoljar’s analysis reveals no connection between what it is like and state- or self-awareness. Further, the ‘for’ emphasized by Nagel drops away as irrelevant. What matters, rather, is that a subject is in a conscious state when that state, because of its nature, makes her feel a certain way. As Stoljar puts it, there is no ‘easy implication’ to a reflexive theory—a theory embracing the transitivity principle. But this does not block the transitivity reading of Nagel’s what it is like phrase. First, it was not claimed that there was an *entailment* from Nagel’s phrase to the transitivity principle. Rather, the transitivity principle provides a good interpretation of the phrase, one revealing its deeper structure. Further, Stoljar’s ‘affective approach’ is flawed. There are plausible cases of states that make us feel ways based on their constitution that nonetheless are not conscious states. If I have an unconscious pain, it may alter my overall mood, it may influence my behavior, etc. precisely because it is a pain and not some other state. But that does not entail that there will be something it is like for me to be in that state. Stoljar’s analysis falls short just where we need it to mark the distinction between conscious and nonconscious states. This suggests a revised analysis: one adding to Stoljar’s characterization *that we are aware of the state*. But this just is a version of the transitivity principle. We can conclude that the principle does not leave out what it is like and so stands as the right way to fix the data a theory of consciousness must explain.

20.2 Varieties of HO Theory

20.2.1 HOP versus HOT

Given the transitivity principle, we can sketch an argument to the best explanation for the HO theory:

1. A mental state is conscious when one is appropriately conscious of oneself as being in that state. (Transitivity Principle)
2. The best explanation of this process is HO representation.
3. Therefore, the best explanation of conscious states is in terms of HO representation.

A key question for HO theory concerns the format of the HO representation. Two main options have been defended in the literature, the higher-order perception (HOP) view (Armstrong, Lycan) and the higher-order thought (HOT) view (Rosenthal, Carruthers; (p. 443) see also Gennaro).¹¹ Armstrong, defending the HOP theory, notes that inner awareness can provide a directed scanning of our experience, in much the way perception provides a directed scanning of our environment. Lycan, too, sees the functional profile of higher-order awareness as analogous to perceptual monitoring of our environment. Further, he contends that the range of features we can be aware of in ex-

perience outstrips conceptual resources, suggesting that the monitor employs perception-like nonconceptual representations.

A key part of Rosenthal's case for HOT theory involves undermining HOP. He argues that perceptual processes are marked by distinctive sensory qualities: mental colors, tastes, smells, etc. However, when we consider conscious experience, we only find the qualities of first-order states. We never experience any higher-order sensory qualities. This suggests that HO representation lacks sensory quality and thus is better viewed as thought-like. In response to Lycan's worry that experience outstrips conceptual resources, Rosenthal argues that this underestimates the range of conceptual resources. If we include comparative concepts like 'lighter than' or 'darker than' we can conceptually represent a vast range of colors, for example, even if we lack distinct concepts like 'red' or 'green' for each quality experience. Further, thought, too, can serve to monitor a domain, as when we monitor our thought for error or bias. Note, however, that since all three theorists are functionalists and representationalists, broadly construed, in the final analysis the difference between HOP and HOT may not be so great.¹²

Another dimension along which HO theorists differ is the question of whether the HO state must be occurrent or whether it can be merely dispositional. Peter Carruthers (2000), in defending a dispositional version of the HO theory, contends that it is implausible that a HO state re-represent all of the detail present in a conscious first-order perceptual state. For one, the information is already present in the first-order state. Why would it need to be re-represented? What evolutionary purpose would that fulfill? And such complete recapitulation at the HO level may well threaten cognitive overload. How could a higher-order cognitive system take up all that data? Instead, Carruthers argues, we should view the presence of the HO state as dispositional. First-order states are conscious when they can be accessed by a higher-order 'mind reading' system. But they need not be occurrently taken up by the system, nor need it re-represent them in all detail.

However, an 'occurrent' HO theorist can respond with several points. First, it is not clear that at any given moment of experience we are actually aware of all that much detail. We can take in a range of properties 'at a glance' by representing them in grouped 'ensembles'.¹³ For example, we can be aware of a pile of colored books in the periphery of our vision without being able to make out the details of the colors and shapes. The actual burden on HO representation at a given time is not so great. Further, there is evidence (p. 444) that we systematically overestimate just how much we can see at a given time. People are surprised to find just how little detail is actively present in parafoveal vision when we are not attending to periphery.¹⁴ This suggests that we may not actively represent many details until we need to, lessening the cognitive burden on HO states. What is more, there are problems with the dispositional approach. One worry is that we are disposed to become aware of many states that do not in fact enter awareness at a given time. I may be disposed to become aware of some sensation of mine if I am prompted to attend to it. For example, can you feel your big toe right now? If the feeling of your big toe was not conscious prior to the prompt, one is nonetheless disposed to become aware of it in the right circumstances. We are left wanting to know which dispositions matter and why. But even if that challenge can be met, it remains unclear that a disposition to represent is enough to make us aware of being in a state. If there is no active, occurrent change in the state or in our awareness of it beyond the potential to represent it, how can that account for the change from a nonconscious state to a conscious one? Thus, it seems an occurrent account is to be preferred.

20.2.2 HOT Theory Detailed

I will now provide a brief sketch of David Rosenthal's HOT theory, to show in more detail how the HO view might be developed.¹⁵ Rosenthal argues that a mental state is conscious when we form an appropriate HOT about

the state. The HOT must possess an assertoric attitude, as other attitudes like wondering or doubting need not make us aware of anything. Further, the state must be formed without any inference or observation the subject is aware of—it must seem spontaneous and immediate from the subject’s point of view. Rosenthal contends that the HOT employs descriptive, conceptual content to represent the target state. Recall that conceptual thought may include comparative concepts to extend the range of conceptual description. Finally, the HOT must make reference to the self. This delivers the following characterization:

HOT: A mental state of seeing a red ball, for example, is conscious when we token the following roughly co-occurrent assertoric HOT: ‘I, myself, am seeing a red ball.’

In addition to picking out intentional contents like ‘ball’, HOTs must be able to pick out the distinctive qualities of sensations. To do so, Rosenthal appeals to ‘quality space’ theory, one characterizing qualities in relational terms, by way of their role in similarity and difference judgments.¹⁶ Further, note that the HOT must be able to pick out mental states as such: I am in a state of seeing or hearing, etc. It may at this point seem that the HOT view cannot be right, as we are never aware of such complex cognitive states when we have ordinary conscious experience. However, according to Rosenthal’s view, HOTs (p. 445) are generally *nonconscious*, so we will not ordinarily be aware of their presence.¹⁷ This also serves to block a potential regress of conscious states making us aware of our conscious states. HOTs, on Rosenthal’s view, are posited to explain the awareness we have of our conscious states. They are not justified on phenomenological grounds, at least not directly. The end result is a nonconscious HOT, formed in a seemingly spontaneous manner, to the effect that ‘I, myself, am seeing a red ball.’ When this HO state is tokened, I become appropriately aware of myself as seeing a red ball—my visual state of seeing a red ball becomes conscious. Armstrong and Lycan offer variations on this sort of theoretical structure, but the broad strokes are the same: a nonconscious first-order state is represented by a separate HO representation, rendering the first-order state conscious.

20.3 Objections to the HO theory

20.3.1 Rock States and Liver States

The first objection we will consider challenges the idea that mere representation is enough to explain the transition from unconscious to conscious mental state. Why is it that when one is conscious of a mental state it becomes conscious, but when one is conscious of a rock, say, it does not? Indeed, we can represent a seemingly unlimited number of things without those things themselves becoming conscious rocks, conscious trees, conscious eggplants, and what have you. If representation does not produce this sort of transformation in the usual case, why think it occurs in the case of mental states? And if more is needed to explain the transition, this suggests that the real work of consciousness is being done by the additional factors, rendering the representational explanation irrelevant. This objection is known as ‘the problem of the rock’.¹⁸

In response, Rosenthal and Lycan argue that the key difference is that mental states are being represented—it is in their nature that when we are conscious of them, they become conscious. But this answer, even if true, needs more support. What is it about the nature of mental states that explains this difference? Here, the HO theorist can point to the self-ascribing character of the HO state. It is an awareness that *I, myself*, am in such-and-such a state. I can be aware of myself as being in a visual state or a state of desire, but it is not clear how I could be aware of myself as being in a rock state (or as being a rock, etc.). Self-ascription thus limits the range of HO awareness. Further, the

way we are built may limit what can trigger the HO representational system. Mental states are causally poised to trigger this system; rocks, trees, and eggplants are not. Even though the HO system employs representation, and thus in principle could represent these other things, the way the system is actually instantiated in us limits its application.

(p. 446) And this helps deal with a variation of the rock objection, the ‘liver state objection’.¹⁹ It does seem that we might self-ascribe states of our own livers: ‘I, myself, am metabolizing bilirubin slowly’, if I have jaundice, say. But such states do not ever become conscious states. Again, it might be that such states are not in a position to trigger the HO system. We are just not built that way. Or perhaps we can form thoughts that we are in such states, but not in a seemingly immediate way. To generate a thought in the spontaneous manner required by the transitivity principle, the right triggering conditions may need to be present. Liver states plausibly fail to meet those conditions. Thus, although liver states may be self-ascribable, they are not properly placed to spontaneously trigger the HO awareness required by the transitivity principle.

20.3.2 Over-sophisticated HOTs

A different set of objections argues that the HOT theory in particular requires a high degree of conceptual sophistication and this in turn rules out many creatures, including perhaps babies or young children, from having conscious mental states.²⁰ The HOT theory demands that we employ concepts picking out mental states as such, as states of seeing, hearing, feeling, and so on. What is more, the requirement that HOTs involve self-representation further limits the creatures that might have conscious states. It may be that most nonhuman animals and children up to three years old lack such conceptual sophistication. But it seems highly implausible that children and animals lack conscious pains or perceptions. Perhaps the HO theorist can bight the bullet here (see Carruthers 1989), but a better tack challenges the claim that the concepts required for HOT are really so sophisticated. Rosenthal contends that nonlinguistic creatures might refer in a minimal way to their sensory states by way of their position in the relevant sensory field. Further, there is evidence of the presence of ‘theory of mind’ capacities in a range of nonhuman animals, including ravens.²¹ Such capacities implicate the employment of an implicit appearance/reality distinction. And this in turn may provide the resources needed by HOT to pick out mental states. So it is not clear that the HOT theory denies consciousness to nonhuman animals and babies.

Further, there is similar evidence of at least a minimal self-concept present in nonhuman animals.²² Social animals in particular need to have ways to track themselves and their conspecifics in terms of their social relations. And an even more basic ability to distinguish one’s own body from the external world is clearly present far down the phylogenetic scale. Are such minimal self-concepts sufficient to explain the sort of self-reference present in consciousness as characterized by the transitivity principle? It appears so, especially when we distinguish between the sort of implicit self-reference present in consciousness and a more full-blown conscious *introspective* awareness of self, present more rarely in us. This sort of introspective self-awareness perhaps does require cognitive (p. 447) resources beyond that of young children. But that is not the kind of self-awareness at issue. Rather, all that is required is the background self-awareness marking our mental states as ours, as integrated into our background awareness of ourselves. This kind of awareness distinguishes my states from the states of others and it provides a background sense of ‘for me-ness’, the phenomenal impression that I subjectively own my conscious states.²³ This more restricted kind of self-awareness is plausibly present in babies and some nonhuman animals, suggesting that the HO theory is not overly restrictive in the way indicated by the objection. The issue is largely empirical, but it is not clear that the HO theory entails an objectionable result.

20.3.3 Misrepresentation and Empty HOTs

A third objection focuses on the representational connection between conscious mental states and our awareness of them. This objection has received the most attention in recent literature, so I will consider it in some detail. The HO theory holds that we represent our conscious states. But by its nature, representation can go astray—successful representation plausibly requires the possibility of misrepresentation.²⁴ So what happens if an HO state misrepresents its target? In particular, what happens to phenomenal character, to ‘what it’s like’ for the subject? Consider the following three scenarios:

1. A subject accurately represents herself as being in a red visual state, by way of an HO representation to the effect that ‘I, myself, am seeing red’, when she is seeing red. (Accurate)
2. A subject inaccurately represents herself as being in a red visual state, by way of an HO representation to the effect that ‘I, myself, am seeing red’, when she is in fact seeing green. (Inaccurate)
3. A subject inaccurately represents herself as being in a red visual state, by way of an HO representation to the effect that ‘I, myself, am seeing red’, when she is in fact in no visual state whatsoever. (Empty)

We can call these scenarios ‘accurate’, ‘inaccurate’, and ‘empty’ scenarios, respectively. What, according to the HO theory, is it like for the subject in each of these cases? In the accurate scenario, it is like seeing red—this is the standard case the theory is designed to explain. But in the inaccurate case, what it is like for the subject is arguably just the same as in the accurate scenario: what it is like for a subject is plausibly a matter of how that subject represents her state. And, by parity of reasoning, what it is like for the subject in the empty case is again the same: it is like seeing red. These results are seen as having devastating consequences for the theory.

First, it seems that the first-order state has been rendered unnecessary for the presence of a conscious state. Conscious experience indistinguishable from the accurate (p. 448) case can occur without a first-order state.²⁵ The relational structure posited to explain state consciousness has collapsed. We are left with a single state accounting for experience, a view with no obvious structural advantage over rival ‘first-order representational’ theories.²⁶ And how could a mere HO state, particularly a nonsensory state like a HOT, explain the richness of conscious experience, in the total absence of the relevant sensation?²⁷

But more seriously, there seems to be an incoherence exposed at the heart of HO theory if the empty scenario can occur. The transitivity principle defines a conscious state as one we are conscious of. In the empty scenario, the state we are conscious of is a red visual state. But that state does not exist. But surely to be a conscious state, the state must exist. So there is no conscious state in the empty scenario. Yet the HO theory holds that what it is like in the empty scenario is subjectively indistinguishable from the accurate case. And if something is subjectively indistinguishable from a conscious experience, it too must be a conscious experience. So in the empty scenario we are in a conscious state! Thus, the empty scenario seems to force the HO theory into an out-and-out contradiction, not a happy result.²⁸ Perhaps the HO theory can hold that it is the HO state itself that is conscious in all cases, but that does not seem to provide an effective explanation of the transitivity principle, of how we are conscious of our conscious states.²⁹ We are not generally aware of our HO states and this is indeed important in blocking a potential regress of conscious states in the theory (see Section 20.2.2 above). And if HO states are conscious for some other reason, we have lost the explanatory power of the theory. We can no longer explain the intransitive consciousness of mental states in terms of HO representation. We are left without an explanation at all. Taken altogether, HO misrepresentation appears to badly undermine the HO theory.

What, then, can the HO theory offer in response? One move is to argue that, at least for the more serious empty case, the lack of a first-order target would block the stable formation of an HO representation. The brain is filled with backward-projecting connections, connections from the higher back to the lower levels. These plausibly create feedback loops which sustain the co-activation and coordination of brain states. It may be that the HO state is connected in this manner to the lower-order areas it monitors. In the empty scenario, the lack of appropriate feedback from the lower-level might block the activity of the targetless HO state. Thus, for architectural reasons we can conclude the empty scenario could not occur. And it might even be that a severe enough mismatch between the HO state and its first-order target would block sustained formation of the HO state, limiting the damage of the inaccurate case.³⁰ While it may seem counterintuitive that our experience of green could be as of red, it is not so implausible that our experience of green might be misrepresented as turquoise or aquamarine. If normal (p. 449) brain functioning serves to limit the possibility of the troubling phenomena, perhaps the HO theory can avoid the objection.

But this sort of response may seem to miss the point. We want to know what accounts for what it is like for the subject according to the HO theory. If it is a matter of how the HO state represents things, we can reasonably ask about cases of misrepresentation even if they are limited in practice (or even nomologically impossible). This is a question of how the theory explains things even in the accurate case. If the HO theory hopes to reductively explain state consciousness in terms of HO representation, we need to know how HO representation accounts for what it is like. And the answer seems to be that it is the content of the HO state alone that explains what it is like. The misrepresentation objection calls out this claim for scrutiny.

There is a fork in the road for HO theorists here. Either they can try to strengthen the relational link between HO and first-order states or they can embrace the idea that HO content is fully sufficient for an explanation of what it is like. The first move leads to either a noncausal (and so, plausibly, nonreductive) view of representation, or it leads to a 'self-representational' theory, where a conscious states represents both itself and the world. Either move will block the empty scenario, but they incur the costs of explaining the strengthened notion of representation or the new self-representational relation. And there is the further problem of restricting the inaccurate scenario, if that worry retains its force in the absence of the empty case. I will not pursue these alternatives further here, though see Chapter 21 on the self-representational approach.³¹

Instead, I will pursue the second option, the claim that HO content is sufficient to fully explain what it is like. There are two related ways of defending this move. First, note that the content of the HO state makes reference to the self: 'I, myself, am in such-and-such a state.' Whenever a subject tokens such a state, the state will successfully refer to that very subject. So there is a sense in which an 'empty' HO state is never really empty. It is always about the very subject thinking that thought. And it then ascribes (in this case erroneously) properties to that subject: that she has the property of being in this or that state. What is more, there will always be various states present in a self-representing subject to serve as an inaccurately-described target. It is not that there is nothing else going on in the mind of the subject when she misattributes a state to herself. And at a minimum, the HOT itself will be present, though misdescribed. So, by the structure of the theory, the empty scenario is impossible: the subject and some state of hers will always be present whenever she tokens a HO state, even if she is inaccurate about which state she is in. And this is not for contingent reasons of neural architecture but because of the nature of the HO states posited.

But we might take a more radical tack. Some HO theorists have defended the idea that in all cases, accurate and inaccurate, the target of the HO state is merely an intentional object.³² The idea here is that all that is required by

the transitivity principle for a state to be conscious is that we are aware of ourselves as being in it. And that can occur even if (p. 450) the state does not exist. We can in some sense be aware of something that is not there. As Gil Harman (1990) notes, there is a decent sense in which Macbeth is aware of a dagger before him, even though there is no such dagger. This explains how things *seem* to Macbeth. In the empty scenario, we represent ourselves as being in a state we are not actually in. But it will seem to us, because we represent things to ourselves that way, that we are in that very state. And so, we will be aware of the state in the sense that counts. It follows that one can be in a conscious state that does not exist, because all that matters for being in a conscious state is how things seem to one, and this is a matter of how we represent our mental lives to ourselves by way of HO representation.

Two further points need to be stressed to make this claim palatable. First, the counter-intuitive nature of this claim can be explained without denying HO representation. We are generally unaware of the HO state making us aware of our conscious states. So that awareness will seem immediate and direct. And because of this, there will not seem to be any 'space' between us and our conscious experience to allow for error. But that is because we remain unaware of the HO mechanism and its workings. The appearance of immediacy misleads us into an overly strong view about the accuracy of inner awareness. What is more, there is no similarly direct means of noticing error because the only seemingly-direct access we have is by way of HO states. We have no cross-modal check on inner awareness in the way we do with perceptual awareness. There is no direct evidence that could undermine our confidence in the deliverances of consciousness, so we conclude that misrepresentation cannot occur. But that does not mean there cannot be theoretical reasons to allow for error.

And this leads to the second point supporting the intentional object view: as general matter, theory sometimes prompts us to give up intuition. We start with the transitivity principle. This pins down the data to be explained by a theory of consciousness. We then posit HO representation as the best explanation of the transitivity principle: this is a theoretical claim, not an a priori analysis or a claim based on intuition. It turns out that HO representation allows for counterintuitive misrepresentation in the ways specified. But so long as the HO theory is well-supported in general, this gives us reason to doubt the intuitions against misrepresentation. In the same sense that it is counterintuitive, though not false, that the earth is in rapid motion, it may be counterintuitive, though not false, that HO representation explains state consciousness, even though it can misrepresent. Sometimes, theory overrides intuition, so long as we have a plausible story saving the appearances. Misrepresentation, though counterintuitive, is not fatal to the HO theory.

20.4 Empirical Support for the Higher-Order Theory

The objections considered thus far have largely been a matter of conceptual coherence. They are not matters to be decided by direct empirical test. This may raise doubts about (p. 451) the empirical status of the HO theory: is it merely a species of a priori analysis, to be accepted or rejected based on imagined counterexamples? Though much of the literature on HO theory focuses on conceptual matters, this is not the case. While the transitivity principle is a matter of pretheoretical analysis, the HO theory is offered as the best explanation of the transitivity principle, one ultimately open to empirical confirmation or refutation. Scientific theories may be challenged both on conceptual and empirical grounds, so it is not a mark against the empirical credentials of the HO view that debate at this stage focuses largely on possible conceptual coherence. Considerable further clarification and refinement will be needed, no doubt, to connect so-called philosophical theories of consciousness with experimental data powerful enough to

separate rival views. Still, even at this stage, we can consider how data from psychology and neuroscience meshes with the HO approach.

The HO theory is committed to the claim that the state accounting for what it is like for us is separate and ‘up-stream’ from our sensory systems. For example, the view appears committed to the idea that activity in primary visual areas alone is insufficient for conscious experience. Rather, there must be activity either in temporal or frontal regions, regions which might realize the sort of metacognitive state required for HO representation. Areas in prefrontal cortex associated with ‘theory of mind’ representations are a reasonable candidate for HO states,³³ as are regions in the dorsolateral prefrontal cortex associated with monitoring conflicting activity at lower levels. Work by Lau and Passingham involving a disparity between accuracy in a perceptual task and judgments of confidence concerning perception implicate activity in dlPFC as corresponding to HO monitoring.³⁴ Lau and Rosenthal argue that these results (and related results) provide evidence for HO theory.³⁵ In addition, Lau and Brown exploit the possibility of HO ‘empty’ misrepresentation to argue that phenomenal illusions like those in Charles Bonnet syndrome offer evidence of consciousness accounted for solely by the presence of the HO state.³⁶ This in effect turns the ‘bug’ of empty representation into an evidential feature. Such claims are controversial and theorists opposed to HO theory have alternative ways of interpreting this data. Still, it seems clear that empirical evidence is relevant to the defense of HO theory. It is not just old-fashioned armchair analysis.³⁷

What is more, there are existing empirical views which may map relatively well to the HO view, broadly considered. The work of Antonio Damasio, for example, involves higher-order self-monitoring mapping activity in bodily and sensory systems.³⁸ And the ‘higher-order syntactic’ view of Edmond Rolls makes explicit connection to HO theory.³⁹ Finally, the work of Hans Flohr offers a way that HO representation might be realized when the right sorts of activation patterns are present in neurons possessing NMDA receptors.⁴⁰ This research has its basis in considerations of the effects of various anesthetics, including the breakdowns in experience that occur under the influence of (p. 452) ketamine. While it is clear that these views can be developed independently of HO theory, the authors have generally recognized the connection and some see it as providing a psychological explanation of the neurological phenomena they have been considering. This further strengthens HO theory’s connection to empirical data, and it recommends the theory as an alternative to the lower-order re-entrant views and global workspace theories popular in neuroscience.⁴¹

20.5 The Hard Problem of Consciousness

There remains, however, an outstanding worry, even if the earlier objections can be defused. Philosophers who hold that no reductive physicalist theory can close the ‘explanatory gap’ between the physical and the phenomenal, or those who think no physicalist theory can handle the ‘hard problem’ of consciousness—the problem of saying why any physical state is conscious at all—will conclude that HO theory falls short in explaining consciousness. It provides no narrowing of the explanatory gap and it fails to convincingly rule out the possibility of zombies, beings physically identical to us that nonetheless lack conscious states. It seems that zombies can have HO states of the relevant kind without being conscious. The view thus fails to explain consciousness in the deep sense required by the gap and the hard problem.

Since this is a challenge that hits all reductive physicalist views, the HO theorist might argue that more general considerations about modality and explanation are the proper place to look for answers to the gap and the hard problem. The conceivability of zombies and the lack of an a priori deduction of the phenomenal facts from the

physical facts may in the end prove irrelevant to the prospects of a successful theory of consciousness.⁴² Still, the HO theory has certain advantages in blocking these objections. First, the HO theory is committed to the claim that the transitivity principle fixes the data a theory of consciousness must explain. This in itself may counter the gap and hard problem. If being in a conscious state is simply being appropriately aware of yourself as being in a state, then the gap falls away. How do we explain consciousness in physical terms? We explain how physically-instantiated higher-order representation makes us aware of our mental states. That is what consciousness amounts to—there is no further question about how that could be, assuming higher-order representation provides an explanation of the appropriate awareness. Further, if some creature represents its first-order states in this way, then it, too, is in conscious states. That is, zombies are inconceivable on the HO theory, because all there is to consciousness is the right sort of awareness of our states and purported zombies can possess this sort of awareness. There is nothing more to a consciousness, so characterized. We have already seen that one can challenge the (p. 453) transitivity principle's claim to adequately fix the data (see Section 20.1.2 above), so I will not revisit the issue here. Suffice to say that *if* one accepts the transitivity characterization, then the gap and the hard problem dissolve.

But it might be argued that the strong intuition of a gap or of the conceivability of zombies remains, even if one accepts the transitivity principle and the positing of HO representation. Indeed, it could be claimed that the transitivity principle is but a necessary condition for consciousness, but it is not sufficient, as the persistence of gap intuitions shows. It may be countered that we should therefore discount the intuitions in the face of a successful theory. But a stronger move is available to the HO theory. The theory itself plausibly predicts the intuitions of an explanatory gap and zombies. And this helps weaken the relevance of those intuitions—if they are to be expected even if the HO theory is true, they lose their undermining force.

So what, according to the HO theory, accounts for the robust presence of these intuitions? First, the mechanism of awareness posited by the theory is generally hidden from first-person access. We are not usually conscious of ourselves as being in higher-order states during conscious experience. They operate unconsciously, behind the scenes. Further, we are unaware of any triggering conditions for the relevant HO states. The awareness they engender appears immediate and noninferential. This means that subjects will not find it intuitive that all conscious states are states represented by HO representation. Thus, we will be surprised to learn of its presence. Further, as noted in discussing intuitions about misrepresentation, the hidden nature of the HO state, as well as the lack of any accessible mediating factor, will support the intuition that we cannot be wrong about what conscious state we are in. There will not seem to be any of the distance or intervening process that explains error in the ordinary perceptual or cognitive case. It will seem to us that we have got direct and incorrigible access to our conscious experience and any claim that there is further background machinery accounting for such awareness will seem counterintuitive. There will thus seem to be a gap between the physically-realized processes posited by the HO theory and phenomenal consciousness. And it will be easy to imagine conscious beings who lack HO representation, because it is counterintuitive that we have such machinery, even if we in fact do.

But what of the distinctive qualities of experience, the 'qualia' supposedly lacking in zombies, unknown by color-deprived super-scientist Mary, etc.? Even if we can explain our seemingly direct access to these things, the things so accessed seem problematic in and of themselves. Qualia have seemingly simple natures: they do not seem to allow decomposition into something more basic, nor do they seem amenable to reductive explanation. They form the basic building blocks of experience and lack any apparent structure to provide the needed 'hooks' for explanation. Further, there is something indescribable or 'ineffable' about qualia. If you have never had an experience with certain basic qualities, there is little informative about the qualities that can be conveyed to you. If you have never tasted sweetness or seen red, for example, no amount of explanation will provide you with what you are missing.

And taken in conjunction with the seeming directness of our first-person access to qualia, they seem only contingently connected to structural, functional, or dynamical properties, properties amenable to (p. 454) physicalist reduction. Directly-accessed qualia are what make consciousness really intractable, it seems.

Again, however, the HO theory can explain why qualia appear intractable when they are not. A guiding principle of cognitive science, one gaining further support in experimental and social psychology, is that the bulk of mental processing occurs outside of conscious awareness.⁴³ We monitor and process a range of stimuli at a given time, but only a select set of features makes it into experience. What is more, there is evidence that unconscious processing is more information-rich, and that consciousness effects a kind of bottleneck in the cognitive stream.⁴⁴ Consciousness seems to be a limited-bandwidth system, one that only takes up features as needed.⁴⁵ Taking these facts into account, it follows that HO representation would have a compressing, limiting effect on the flow of information. There would be pressures to reduce to a minimum the information needed to make the subject aware of relevant features of her conscious states. It is plausible, then, that the HO state would abstract away from the range of causal and functional connections constitutive of physically-reducible sensory qualities. All that HO states would register would be a compressed version of the information, that the state possesses this or that quality, without making explicit the connections to other features of the mind. Or a complex battery of associations, exemplars, and contextual cues might be reduced down to a representation of a single, constant quality, the one most likely to track salient and useful features. Doubtless, this compressing and limiting process occurs throughout the perceptual processing stream, but awareness of our states filtered through HO representation would add to this effect.

The result of this ongoing compressing and limiting would be the appearance of seemingly simple qualities, qualities without the sorts of relational connections needed to provide an informative description. And, for that reason, we would be unaware of connections between the qualities and any structural, functional, or dynamical properties of mental states—they would seem only contingently connected to such things. We would have the appearance of simple, indescribable qualities, only contingently connected to the mind. Further, because these qualities are known in a seemingly direct manner, our knowledge of them would seem special and privileged. We would find it counterintuitive that we could be in error about them. There would therefore seem to be an unbridgeable explanatory gap and zombies would be easy to conceive. But because this is predicted by the constraints of architecture and resources imposed by HO theory, we can reasonably discount the importance of these intuitions. They are just what we would expect if the HO theory is true. And this further vindicates the claim made in Section 20.1.1 above that the transitivity is a full and adequate characterization of the data a theory must explain. It predicts that we would find a phenomenal ‘what it’s like’ characterization appealing, even though the transitivity principle captures the data without remainder. Thus, the HO theory disarms the gap and the hard problem, clearing the way for a satisfying explanation of consciousness in physical terms.

References

Allen, C. and Bekoff, M. (1997), *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. Cambridge, MA: MIT Press.

[+] Find this resource:

Armstrong, D. M. (1968), *A Materialist Theory of Mind*. London: Routledge and Kegan Paul.

[+] Find this resource:

Armstrong, D. M. (1981), ‘What is Consciousness?’ in *The Nature of Mind*. Ithaca, NY: Cornell University Press.

[+] Find this resource:

Bekoff, M. (2003), *Minding Animals: Awareness, Emotions, and Heart*. Oxford: Oxford University Press.

[+] Find this resource:

Block, N. (1995), 'On a Confusion about the Function of Consciousness', *Behavioral and Brain Sciences*, 18: 227–87.

[+] Find this resource:

Block, N. and Stalnaker, R. (1999), 'Conceptual Analysis, Dualism, and the Explanatory Gap', *Philosophical Review*, 108: 1–46.

[+] Find this resource:

Braddon-Mitchell, D. and Jackson, F. (2007), *Philosophy of Mind and Cognition*, 2nd edn. Oxford: Blackwell Publishing.

[+] Find this resource:

Brown, R. (2015), 'The HOROR Theory of Phenomenal Consciousness', *Philosophical Studies*, 172/7: 1783–94.

[+] Find this resource:

Bugnyar, T., Reber, S., and Buckner, C. (2016), 'Ravens attribute visual access to unseen competitors', *Nature Communications*. 7.

[+] Find this resource:

Byrne, A. (1997), 'Some like it HOT: Consciousness and higher-order thoughts', *Philosophical Studies*, 2/2:103–29.

[+] Find this resource:

Byrne, Alex (2004), 'What phenomenal consciousness is like' in R. J. Gennaro (ed.), *Higher-order Theories of Consciousness*. Amsterdam: John Benjamins Publishers, 203–25.

[+] Find this resource:

Carruthers, P. (1989), 'Brute experience', *Journal of Philosophy*, 86: 258–69.

[+] Find this resource:

Carruthers, P. (2000), *Phenomenal Consciousness*. Cambridge: Cambridge University Press.

[+] Find this resource:

Carruthers, P. (2005), *Consciousness: Essays from a Higher-Order Perspective*. New York: Oxford University Press.

[+] Find this resource:

Caston, V. (2002), 'Aristotle on consciousness', *Mind*, 111/444: 751–815.

[+] Find this resource:

Chalmers, D. J. (1996), *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

[+] Find this resource:

Chalmers, D. J. (2013), 'How can we construct a science of consciousness?', *Annals of the New York Academy of Sciences*, 1303: 25–35.

[+] Find this resource:

Clark, A. (1993), *Sensory Qualities*. Oxford: Clarendon Press.

[+] Find this resource:

Cohen, M. A., Dennett D. C., and Kanwisher, N. (2016), 'What is the Bandwidth of Perceptual Experience?' *Trends in Cognitive Science*, 20/5: 324–35.

[+] Find this resource:

Damasio, A. (1999), *The Feeling of What Happens*. New York: Houghton Mifflin Harcourt.

[+] Find this resource:

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., and Sergent, C. (2006), 'Conscious, preconscious, and subliminal processing: a testable taxonomy', *Trends in Cognitive Science*, 15/8: 365–73.

[+] Find this resource:

Dretske, F. (1981), *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books/MIT Press.

[+] Find this resource:

Dretske, F. (1993), 'Conscious experience', *Mind*, 102/406: 263–83.

[+] Find this resource:

Dretske, F. (1995), *Naturalizing the Mind*. Cambridge, MA: Bradford Books/MIT Press.

[+] Find this resource:

Flohr, H. (1995), 'Sensations and brain processes', *Behavioral Brain Research*, 71: 157–61.

[+] Find this resource:

Flohr, H. (1999), 'NMDA-receptor-mediated computational processes and phenomenal consciousness' in T. Metzinger (ed.), *Neural Correlates of Consciousness*. Cambridge, MA: MIT Press, 245–58.

[+] Find this resource:

(p. 456) Fodor, J. A. (1990), *A Theory of Content and Other Essays*. Cambridge, MA: Bradford Books/MIT Press.

[+] Find this resource:

Gennaro, R. J. (2002), 'Jean-Paul Sartre and the HOT theory of consciousness', *Canadian Journal of Philosophy*, 32/3: 293–330.

[+] Find this resource:

Gennaro, R. J. (2012), *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. Cambridge, MA: Bradford Books/MIT Press.

[+] Find this resource:

Goldman, A. I. (1993), 'Consciousness, Folk Psychology, and Cognitive Science', *Consciousness and Cognition*, 2: 264–82.

[+] Find this resource:

Güzeldere, G. (1995), 'Is consciousness the perception of what passes in one's own mind?' in T. Metzinger (ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh, 335–57.

[+] Find this resource:

Harman, G. (1990), 'The intrinsic quality of experience', *Philosophical Perspectives*, 4: 31–52.

[+] Find this resource:

Kriegel, U. (2003), 'Consciousness as intransitive self-consciousness: Two views and an argument', *Canadian Journal of Philosophy*, 33: 103–32.

[+] Find this resource:

Kriegel, U. (2009), *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.

[+] Find this resource:

Lamme, V. A. (2006), 'Towards a true neural stance on consciousness', *Trends in Cognitive Science*, 10: 494–501.

[+] Find this resource:

Land, M. F. and Tatler, B. W. (2009), *Looking and Acting: Vision and Eye Movements in Natural Behavior*. New York: Oxford University Press.

[+] Find this resource:

Lau, H. and Brown, R. (2019), 'The Emperor's New Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations', in A. Pautz and D. Stoljar (eds.), *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*. Cambridge, MA: MIT Press, 171–98.

[+] Find this resource:

Lau, H. and Passingham, R. (2006), 'Relative blindsight and the neural correlates of visual consciousness', *Proceedings of the National Academy of Science*, 103: 18763–9.

[+] Find this resource:

Lau, H. and Rosenthal, D. (2011), 'Empirical support for higher-order theories of conscious awareness', *Trends in Cognitive Sciences*, 15/8: 365–73.

[+] Find this resource:

Lewis, D. K. (1972), 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy*, L/3: 249–58.

[+] Find this resource:

Lycan, W. G. (1987), *Consciousness*. Cambridge, MA: MIT Press.

[+] Find this resource:

Lycan, W. G. (1996), *Consciousness and Experience*. Cambridge, MA: MIT Press.

[+] Find this resource:

Lycan, W. G. (2004), 'The Superiority of HOP to HOT' in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins Publishers.

[+] Find this resource:

Millikan, R. (1984), *Language, Thought, and Other Biological Categories*. Cambridge, MA: Bradford Books/MIT Press.

[+] Find this resource:

Nagel, T. (1974), 'What is it like to be a Bat?' *Philosophical Review*, 83: 435–56.

[+] Find this resource:

Neander, K. (1998), 'The Division of Phenomenal Labor: A Problem for Representational Theories of Consciousness', *Noûs*, 32/S12: 411–34.

[+] Find this resource:

Norretranders, T. (1991), *The User Illusion: Cutting Consciousness Down to Size*. Viking Penguin.

[+] Find this resource:

O'Regan, J. K. (2011), *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*. Oxford: Oxford University Press.

[+] Find this resource:

Rolls, E. T. (2004), 'A higher order syntactic thought theory of consciousness' in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins Publishers.

[+] Find this resource:

Rosenthal, D. M. (1986), 'Two concepts of consciousness', *Philosophical Studies*, 49/3: 329–59.

[+] Find this resource:

Rosenthal, D. M. (2000), 'Consciousness and Metacognition' in D. Sperber (ed.), *Metarepresentation: Proceedings of the Tenth Vancouver Cognitive Science Conference*. New York: Oxford University Press.

[+] Find this resource:

(p. 457) Rosenthal, D. M. (2004), 'Varieties of Higher-Order Theory' in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins Publishers, 19–44.

[+] Find this resource:

Rosenthal, D. M. (2005), *Consciousness and Mind*. Oxford: Clarendon Press.

[+] Find this resource:

Rosenthal, D. M. (2011), 'Exaggerated reports: reply to Block', *Analysis*, 71/3: 431–7.

[+] Find this resource:

Saxe, R. (2009), 'Theory of mind: Neural basis' in W. Banks (ed.), *Encyclopedia of Consciousness*. Oxford: Academic.

[+] Find this resource:

Seager, W. (1999), *Theories of Consciousness: An Introduction and Assessment*. New York: Routledge.

[+] Find this resource:

Sebastián, M. Á. (2014), 'Not a HOT Dream' in R. Brown (ed.), *Phenomenology and the Neurophilosophy of Consciousness*, Studies in Brain and Mind, Vol. 6. New York: Springer.

[+] Find this resource:

Siewert, C. (1998), *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.

[+] Find this resource:

Stoljar, D. (2016), 'The Semantics of "What it's like" and the Nature of Consciousness', *Mind*, 125/500: 1161–98.

[+] Find this resource:

Tye, M. (1995), *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.

[+] Find this resource:

Van Gulick, R. (2004), 'Higher-order global states (HOGS): An alternative higher-order model' in R. J. Gennaro (ed.), *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins Publishers.

[+] Find this resource:

Ward, E. J., Bear, A., and Scholl, B. J. (2016), 'Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access', *Cognition*, 152: 78–86.

[+] Find this resource:

Weisberg, J. (2008), 'Same old, same old: The same-order representation theory of consciousness and the division of phenomenal labor', *Synthese*, 160/2: 61–81.

[+] Find this resource:

Weisberg, J. (2011a), 'Review of *Subjective Consciousness: A Self-Representational Theory*, by Uriah Kriegel', *Mind*, 120/478: 538–42.

[+] Find this resource:

Weisberg, J. (2011b), 'Misrepresenting Consciousness', *Philosophical Studies*, 154: 409–33.

[+] Find this resource:

Weisberg, J. (2011c), 'Abusing the notion of "what-it's-like-ness": A reply to Block', *Analysis*, 71/3: 438–43.

[+] Find this resource:

Weisberg, J. (2014a), 'Review of *The Consciousness Paradox: Consciousness Concepts, and Higher-Order Thoughts*, by Rocco J. Gennaro', *Australasian Journal of Philosophy*, 92/2: 401–4.

[+] Find this resource:

Weisberg, J. (2014b), 'Sweet Dreams are Made of This? A HOT Response to Sebastián' in R. Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, Studies in Brain and Mind, Vol. 6. (New York: Springer), 433–43.

[+] Find this resource:

Wilberg, J. (2010), 'Consciousness and false HOTs', *Philosophical Psychology*, 23/5: 617–38.

[+] Find this resource:

Wilson, T. D. (2002), *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

[+] Find this resource:

Zahavi, D. and Kriegel, U. (2015), 'For-me-ness: What it is and what it is not' in D. Dahlstrom, A. Elpidorou, and W. Hopp (eds), *Philosophy of Mind and Phenomenology*. Abingdon: Routledge, 36–53.

[+] Find this resource:

Notes:

(¹) Though there are nonreductive versions of the HO theory (see for example Chalmers 2013), in this chapter I will focus only on those arguing for a reduction of consciousness to HO representation.

(²) Another use of 'consciousness' picks out the consciousness of creatures: a creature can be conscious or unconscious. The HO theory focuses on 'state consciousness' and leaves 'creature consciousness' to one side. Cf. Rosenthal 1986.

(³) For example, Fodor 1990; Dretske 1981; Millikan 1984.

(⁴) In what follows, I will use the terms 'aware of' and 'conscious of' interchangeably, and I intend them to pick out transitive consciousness characterized independently of intransitive state consciousness.

(⁵) Lewis 1972; Cf. Braddon-Mitchell and Jackson 2007.

(⁶) See Zahavi and Kriegel 2015.

(⁷) See, e.g., Caston 2002; Güzeldere 1995; Gennaro 2002.

(⁸) See also Byrne 2004.

(⁹) See Seager 1999: ch 3, for a detailed discussion of this case.

(¹⁰) This kind of consciousness is known as 'phenomenal consciousness.' See Block 1995; Chalmers 1996: ch 1.

(¹¹) Armstrong 1968, 1981; Lycan 1987, 1996; Rosenthal 1986, 2005; Carruthers 2000, 2005; Gennaro 2012.

(¹²) See Rosenthal 2004; Lycan 2004, for further discussion of the differences between HOP and HOT theories.

(¹³) Cohen et al 2016; Ward et al 2016.

(¹⁴) See Land and Tatler 2009: ch 3.

(¹⁵) See Rosenthal 2005.

(¹⁶) Rosenthal 2005: chs 5–7; Cf. Clark 1993.

(¹⁷) HOTs themselves can become conscious when targeted by further '3rd-order' states. Rosenthal labels this phenomenon 'introspection'.

(¹⁸) Goldman 1993.

(¹⁹) Block 1995.

(²⁰) Block 1995: 280.

(²¹) Bugnyar et al 2016; see also Allen and Bekoff 1997.

(²²) Bekoff 2003.

(²³) See Zahavi and Kriegel 2015.

(²⁴) Dretske 1981.

(²⁵) See Wilberg 2010 for a challenge to this claim. See also Gennaro 2012, 59–66.

(²⁶) Dretske 1995 and Tye 1995 defend first-order representational theories. See Byrne 1997 for details of this objection.

(²⁷) See Byrne 1997; Neander 1998 for extended treatments of the objection. See also Van Gulick 2004.

(²⁸) Kriegel 2003. Cf. Kriegel 2009.

(²⁹) Though see Brown 2015.

(³⁰) See Gennaro 2012: ch 9, for elaboration and defense of these claims.

(³¹) See, for example, Kriegel 2009. See Weisberg 2008, 2011a, 2014a for criticisms of this approach.

(³²) Rosenthal 2004, 2011; Weisberg 2011b, 2011c.

(³³) Medial PFC, for example; see Saxe 2009.

(³⁴) Lau and Passingham 2006.

(³⁵) Lau and Rosenthal 2011.

(³⁶) Lau and Brown 2019.

(³⁷) See Sebastián 2014 for an interesting empirical challenge to HO theory. See Weisberg 2014b for an HO response.

(³⁸) Damasio 1999.

(³⁹) Rolls 2004.

(⁴⁰) Flohr 1995, 1999.

(⁴¹) See Dehaene et al 2006; Lamme 2006.

(⁴²) See, e.g., Block and Stalnaker 1999.

(⁴³) See, for example, Wilson 2002.

(⁴⁴) Norretranders 1991.

(⁴⁵) See O'Regan 2011; Cf. Cohen et al 2016.

Josh Weisberg

Josh Weisberg is associate professor in philosophy at the University of Houston. He is the author of *Consciousness* (Polity 2014).

PRINTED FROM OXFORD HANDBOOKS ONLINE (www.oxfordhandbooks.com). © Oxford University Press, 2018. All Rights Reserved. Under the terms of a title in Oxford Handbooks Online for personal use (for details see Privacy Policy and Legal Notice).

Subscriber: null; date: 09 March 2021