HIGHER-ORDER THEORIES OF CONSCIOUSNESS

David Rosenthal
Philosophy and Cognitive Science
Graduate Center, City University of New York
Email: davidrosenthal1@gmail.com
ORCiD: https://orcid.org/0000-0002-9323-0424

This manuscript was solicited for the MIT Open Encyclopedia of Cognitive Science, and after many months of silence rejected without explanation.

Abstract

Many perceptions, thoughts, and emotions occur consciously. But many others occur without being conscious. Higher-order (HO) theories of consciousness seek to explain how mental states that are conscious differ from those that are not. That would tell us what it is for a mental state to be conscious. All HO theories rest on the commonsense observation that whenever a mental state is conscious, one is in some way aware of being in that state. So they see awareness of being in a mental state as a necessary condition for that state to be conscious. The reliance on this condition distinguishes HO theories from other theories of consciousness. There are several versions of a HO theory, which differ mainly in how they explain our awareness of mental states when those states are conscious.

History

The history of theories of consciousness is not very old, and that holds for HO theories in particular. Prior to the late 19th century, it was widely held that all mental states are conscious. If that were so, one could explain what it is for a state to be conscious simply by specifying what it is for the state to be mental. No additional explanation would be needed.

So prior to the late 19th century there was little discussion of the consciousness of mental states, and almost no theorizing. There was ample discussion of introspective awareness, but that was rarely tied to mental states' being conscious. One salient exception is John Locke's remark that "[c]onsciousness is the perception of what passes in a Man's own mind" (1975, book II, ch. I, §19, p. 115). But Locke regarded all mental states as conscious, and did not elaborate on that idea. So that lone remark is not an early version of a HO theory. By contrast, Gottfried Wilhelm Leibniz (2014) did explicitly countenance the possibility of unconscious perceptions, and so recognized that being conscious is different from being mental. And he posited apperception as a distinct mental factor that results in some mental states' being conscious. This is likely the earliest statement of a genuine HO approach to consciousness.

Increasing recognition in the late 19th century that many mental states do occur unconsciously led many to try to explain what it is for a mental state to be conscious, as against simply being mental. The first developed HO theory likely occurs in Friedrich Nietzsche (2001; see Riccardi 2021). But the sustained history of HO theories begins in the mid-20th century with David Armstrong (1968), David Rosenthal (1986), and Peter Carruthers (1996), each of whom independently developed versions of the theory.

Core concepts

As noted above, all HO theories rely on the commonsense observation that a mental state is conscious only if one is in some way aware of being in that state. That observation is widely known as the transitivity principle (Rosenthal 1997). It is hard to deny, since we would not count any mental state as conscious if the individual were wholly unaware of being in the state. But the transitivity principle is best seen not as part of any theory of consciousness, but rather as a pretheoretic way of specifying what consciousness is. The work of a HO theory, then, is to explain the nature of the HO awareness that is necessary for a mental state to be conscious.

Almost all HO theories hold that such HO awareness consists of a HO mental state that is distinct from the first-order mental state one is aware of. And most HO theories also hold that such HO states need not themselves be conscious. Indeed, for the HO state to be conscious would, on most HO theories, require a third-order state that makes one aware of that second-order state, threatening a regress. Because HO states on most HO theories are rarely conscious, those theories do not appeal to introspection to establish that HO states occur. Rather, the HO states are theoretical posits designed to explain what it is for first-order mental states to be conscious.

In contrast with HO theories, first-order theories seek to explain what it is for a mental state to be conscious independently of any awareness of the state. Some first-order theories appeal just to correlations with neurological functioning (e.g., Block 2007), and others to psychological functioning apart from any HO awareness (e.g., Dehaene & Naccache 2001). Still, first-order theories rarely deny that one is in some way aware of any mental state that is conscious (Naccache 2019).

Metacognition is the capacity to monitor and cognitively assess what one knows. Because metacognition operates on first-order cognitive states, it is sometimes compared to the HO awareness posited by HO theories (Proust, 2013. But one can be aware of a cognitive state without being able to assess that state cognitively. And we sometimes unconsciously monitor cognitive states that we are wholly unaware of. So the HO awareness of HO theories is different from metacognition.

Questions, controversies, and new developments

The main controversy that divides HO theories is about the way one is aware of being in a mental state when that state is conscious. One possibility echoes Locke's remark mentioned above, that consciousness consists in perceiving the contents of one's mind. On that type of HO theory, advanced by Armstrong (1968), the HO awareness consists in perceiving those states. The major alternative, pioneered by Rosenthal (1986, 2005), is that when a mental state is conscious one is aware of being in the state by having a HO thought that one is in that state.

The appeal to HO perception can seem highly intuitive. Still, there is no relevant sense modality for HO perception; so a perceptual model of HO awareness may seem little more than a metaphor. One might urge that the neurological implementation of the HO awareness is like an inner sense (Lau, 2022, §7.5). But because consciousness is itself a psychological phenomenon, we would like an explanation cast in psychological terms.

HO thoughts have the advantage that they can specify in highly fine-grained ways exactly how one is subjectively aware of one's first-order states. And though an appeal to HO thoughts can seem overly intellectual, that concern may be largely dispelled by noting that HO thoughts rarely occur consciously. Also, Armstrong's own appeal to HO perception is very close to a HO-thought theory, since he holds that perceiving itself is exclusively a matter of conceptual content.

Mainstream HO theories differ in other respects as well. Armstrong and Rosenthal both posit occurrent HO states that are distinct from the first-order mental states that they make one aware of. Carruthers (1996), by contrast, has posited dispositions for HO thoughts to occur, though it is not obvious how a disposition to have a HO thought could make one aware of being in a mental state. And Uriah Kriegel's (2009) self-representational theory holds that the HO awareness is not distinct from the state it makes one aware of, but part of the very same state. Still, such a single state would have two distinct factors, the HO awareness and the rest of the state. And there are concerns about whether one can sustain holding that these two factors constitute a single state (Phillips, 2014).

Because most HO theories posit a HO awareness that is distinct from the first-order state it makes one aware of, some have objected that the HO state could misrepresent the mental properties of the first-order state, and even that a HO state could make one aware of being in a first-order state that does not occur at all (Levine, 2001, ch. 4; Block 2011). But HO theories do not predict or imply that those things do ever occur. So a HO theory could simply add a stipulation that they never do. Evidently, the real objection is that a theory of consciousness should rule that such occurrences are impossible, though it is not clear why.

In any case, the objection from misrepresentation is arguably misguided. For one thing, it rests on a conception of what it is for a mental state to be conscious that is at best controversial (Weisberg 2011). And that aside, there is compelling evidence that the way individuals are subjectively aware of their mental states is sometimes inaccurate. A dramatic case occurs in change blindness, when one does not consciously see a salient

change in an object (Grimes 1996). One can then be subjectively aware of perceiving the object in respect of its pre-change feature even though there is decisive evidence that the visual system is registering the post-change feature (Fernandez-Duque & Thornton 2000). And being subjectively aware of a mental state that does not occur at all is arguably simply a special case of such HO misrepresentation.

The objection from misrepresentation also ignores an important advantage of any HO theory. Consciousness is a matter of mental appearance. It is the way our mental lives subjectively appear to us. The HO states that HO theories posit explain those mental appearances, which can on occasion diverge from actual psychological functioning. Some have contended that consciousness cannot accommodate a distinction between appearance and reality (Nagel 1974). HO theories dispute that, since a HO state would constitute the reality of consciousness and the content of that HO state would constitute the subjective appearances. In any case, consciousness does not exhaust mental functioning. So we can distinguish the mental appearances of consciousness from the underlying mental reality that occurs independently of consciousness (Rosenthal 2022).

Broader connections

HO theories always posit two factors, a HO factor that explains the subjective appearances, and a first-order factor that explains psychological functioning independently of those appearances. So HO theories are especially useful in explaining occurrences in which subjective awareness diverges from psychological functioning. One example, just noted, is change blindness, in which the way one is subjectively aware of seeing an object can diverge from the way that object affects the visual system. HO theories readily explain such occurrences. It is unclear that any other theory of consciousness has the resources to do so.

There are other examples. In blindsight, damage to primary visual cortex prevents subjective awareness of stimuli even though those stimuli demonstrably affect psychological functioning. Again, HO theories readily explain how this can occur (Weiskrantz 2009). And there are other neurological and psychological considerations that also provide strong support for a HO theory (Lau & Passingham 2006; Lau & Rosenthal 2011).

Further reading

- Gennaro, Rocco J., ed., Higher-Order Theories of Consciousness, Amsterdam and Philadelphia: John Benjamins Publishers, 2004.
- Weisberg, Josh, and David Rosenthal (forthcoming), “Higher-Order Theories of Consciousness,” in Consciousness: A Comprehensive Reference, 2nd edn., ed. Antonino Raffone, Amsterdam: Elsevier.
- Weisberg, Josh (2020), "Higher-Order Theories of Consciousness,” in Oxford Handbook of the Philosophy of Consciousness, ed. Uriah Kriegel, Oxford: Oxford University Press, (2020), pp. 438-457.

References

Armstrong, D. M. (1968), A Materialist Theory of the Mind, New York: Humanities Press; second revised edition, London: Routledge & Kegan Paul, 1993.

Block, Ned (2007), "Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience, Behavioral and Brain Sciences, 30, 5-6 (December): 481-499.

Block, Ned (2011), "The Higher Order Approach to Consciousness is Defunct," Analysis, 71, 3 (July): 419-431.

Carruthers, Peter (1996), Language, Thought and Consciousness: An Essay in Philosophical Psychology, Cambridge: Cambridge University Press.

Dehaene, Stanislas, and Lionel Naccache, (2001), "Towards a Cognitive Neuroscience of Consciousness: Basic' Evidence and a Workspace Framework, Cognition, 79, 1-2 (April), 1-37.

Fernandez-Duque, Diego, and Ian M. Thornton (2000), "Change Detection without Awareness: Do Explicit Reports Underestimate the Representation of Change in the Visual System?", Visual Cognition, 7, 1-2-3 (January–March): 324–344.

Grimes, John (1996), "On the Failure to Detect Changes in Scenes across Saccades," in Perception, ed. Kathleen Akins, New York: Oxford University Press, pp. 89–110.

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

Lau, Hakwan (2022), In Consciousness We Trust, Oxford: Oxford University Press.

Lau, Hakwan, and Richard E. Passingham (2006), "Relative Blindsight in Normal Observers and the Neural Correlate of Visual Consciousness. Proceedings of the National Academy of Sciences, 103, 9 (December 5): 18763-18768.

Lau, Hakwan, and David Rosenthal (2011), "Empirical Support for Higher-order Theories of Conscious Awareness," Trends in Cognitive Sciences, 15, 8 (August): 365-373.

Leibniz, Gottfried Wilhelm (2014), Monadology, tr. Lloyd Strickland, Edinburgh: Edinburgh University Press.

Levine, Joseph (2001), Purple Haze: The Puzzle of Consciousness, New York: Oxford University Press.

Locke, John, An Essay Concerning Human Understanding, edited from the 4th (1700) edn. by Peter H. Nidditch, Oxford: Clarendon Press, 1975.

Naccache, Lionel (2018), "Why and How Access Consciousness Can Account for Phenomenal Consciousness," Philosophical Transactions of the Royal Society B: Biological Sciences, 373, 1755 (September 9): Article 20170357.

Nagel, Thomas (1974), “What Is It Like to Be a Bat?”, The Philosophical Review, 83, 4 (October): 435–50.

Nietzsche, Friedrich (2001), The Gay Science. ed. Bernard Williams, tr. Josefine Nauckhoff and Adrian Del Caro, Cambridge: Cambridge University Press.

Phillips, Ben (2014), “Indirect Representation and the Self-Representational Theory of Consciousness,” Philosophical Studies, 167, 2 (January): 273-290.

Proust, Joëlle (2013), The Philosophy of Metacognition: Mental Agency and Self-Awareness, Oxford: Oxford University Press.

Riccardi, Mattia (2021), Nietzsche’s Philosophical Psychology, Oxford: Oxford University Press.

Rosenthal, David (1986), “Two Concepts of Consciousness,” Philosophical Studies, 49, 3 (May 1986): 329-359.

Rosenthal, David (1997), “A Theory of Consciousness,” in The Nature of Consciousness: Philosophical Debates, ed. Ned Block, Owen Flanagan, and Güven Güzeldere, Cambridge, MA: MIT Press, 1997, pp. 729-753.

Rosenthal, David (2005), Consciousness and Mind, Oxford: Clarendon Press.

Rosenthal, David (2022), “Mental Appearance and Mental Reality,” in Qualitative Consciousness: Themes from the Philosophy of David Rosenthal, ed. Josh Weisberg Cambridge: Cambridge University Press, pp. 243-271.

Weisberg, Josh (2011), “Misrepresenting Consciousness,” Philosophical Studies, 154, 3 (July): 409-433.

Weiskrantz, Lawrence (2009), Blindsight: A Case Study Spanning 35 Years and New Developments, 2nd edition, Oxford: Oxford University Press.