MIT Press • Open Encyclopedia of Cognitive Science

Higher-Order Theories of Consciousness

David Rosenthal[1]

[1]City University of New York

MIT Press

Published on: Apr 22, 2026

URL: https://oecs.mit.edu/pub/qctxcxis

Higher-order theories of consciousness seek to explain what it is for a mental state to be conscious. In particular, they seek to explain how mental states that are conscious differ from mental states that are not. All higher-order theories rest on the commonsense observation that whenever a mental state is conscious, one is in some way aware of being in that state. In adopting that commonsense observation, higher-order theories see one's being aware of being in a mental state as a necessary condition for that state to be conscious. The reliance on this necessary condition distinguishes higher-order theories from other theories of consciousness. There are several versions of higher-order theory, which differ mainly in how they explain one's awareness of a mental state when that state is conscious.

History

The history of theories of consciousness is not very old, and that holds for higher-order theories in particular [see Consciousness]. 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 "consciousness is the perception of what passes in [one's] own mind" (Locke, 1689/1975). 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 higher-order theory.

By contrast, Gottfried Wilhelm Leibniz (1686/2014) did explicitly countenance the possibility of unconscious perceptions. So he recognized that a mental state's being conscious consists in something in addition to those properties in virtue of which the state is a mental state.

And Leibniz posited what he called apperception, in effect a higher-order awareness of a state that is conscious, to explain its being conscious. This is likely the earliest statement of a genuine higher-order 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 higher-order theory likely occurs in Friedrich Nietzsche (1882/2001; see Riccardi, 2021). But the sustained history of higher-order theories begins in the mid-20th century with David Armstrong (1968/1993), David Rosenthal (1986), and Peter Carruthers (1996), each of whom independently developed versions of the theory.

Core concepts

All higher-order 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 common sense does not count any mental state as conscious if the individual is wholly unaware of being in the state. So it is a necessary condition for a mental state to be conscious that one is in some way aware, or conscious, of being in that state.

But the transitivity principle is not best seen as part of any theory of consciousness but rather as a pretheoretic way of specifying what it is for a mental state to be conscious. And because the property of a mental state's being conscious is different from the property of one's being conscious of something, the transitivity principle fixes what it is for a state to be conscious in a noncircular, informative way. The work for a higher-order theory, then, is to explain the nature of the higher-order awareness that is necessary for a mental state to be conscious.

Almost all higher-order theories hold that such higher-order awareness consists of a higher-order mental state that is distinct from the first-order mental state one is aware of. A perception, for example, would be conscious only if it is accompanied by a distinct mental state that makes one aware that one has that perception. And most higher-order theories also hold that those higher-order states need not themselves be conscious. Accordingly, one would be aware of having the perception but that awareness of the perception would not itself be a conscious state.

Indeed, for the higher-order state to be conscious would, on most higher-order theories, require a third-order state that makes one aware of that second-order state, threatening a regress. Because higher-order states on most higher-order theories are rarely conscious, those theories do not appeal to introspection to establish that higher-order states occur. Rather, the higher-order states are theoretical posits designed to explain what it is for first-order mental states to be conscious.

In contrast with higher-order theories, there are so-called first-order theories, which 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 exclusively to correlations with neurological functioning (e.g., Block, 2007). The global workspace theory, which is also a first-order theory, posits that a state is conscious if it is globally available for processing and that availability need not involve any higher-order awareness of the state that is conscious (e.g., Dehaene & Naccache, 2001). Still, such first-order theories sometimes note that one will be in some way aware of any mental state that is conscious (Naccache, 2018).

Metacognition is the capacity to monitor and cognitively assess what one knows [see Metacognition]. Because metacognition operates on first-order cognitive states, it is sometimes compared to the higher-order awareness posited by higher-order theories (Proust, 2013).

But one can be aware of a cognitive state without being able to assess that state cognitively. And a person can sometimes unconsciously monitor cognitive states that the person is wholly unaware of.

A common example of such unconscious monitoring occurs when an individual stops thinking consciously about an unresolved problem but sometime later finds that the solution suddenly consciously occurs to them. Evidently, there must in such a case have been unconscious thinking that led to the solution, whose steps were unconsciously monitored for relevance. The higher-order awareness of higher-order theories is plainly different from metacognition.

Questions, controversies, and new developments

The main controversy that divides higher-order 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 that consciousness consists in perceiving the contents of one's mind. On that perceptual model, the higher-order awareness consists in one's perceiving those states (Armstrong, 1968/1993). The major alternative is the higher-order-thought theory, in which one is aware of being in a mental state that is conscious by having a higher-order thought that one is in that state (Rosenthal, 1986, 2005).

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

Higher-order 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 higher-order thoughts can seem overly intellectual, that concern may be largely dispelled by noting that higher-order thoughts rarely occur consciously. Also, the higher-order perception theory is often very close to a higher-order-thought theory, since proponents of the higher-order perception theory often hold that perceiving is itself simply a matter of conceptual content (Armstrong, 1968/1993).

Mainstream higher-order theories differ in other respects as well. Armstrong (1968/1993) and Rosenthal (1986) both posit occurrent higher-order states that are distinct from the first-order mental states that they make one aware of.

In contrast, there is also a dispositional higher-order-thought theory, on which a state is conscious just in case one is disposed to have a higher-order thought that one is in that state (Carruthers, 1996). One need not on this theory actually have a higher-order thought, whether conscious or not, so long as one is disposed to have one. But it is not obvious how being disposed to have a higher-order thought could make one aware of being in a mental state.

There is also a self-representational theory, in which the higher-order awareness is not distinct from the state it makes one aware of but part of the very same state (Kriegel, 2009). Still, such a single state would nonetheless have two distinct factors, the higher-order awareness and the rest of the state, so it may be unclear what the theoretical significance is of maintaining that the higher-order awareness is part of the

state it makes one aware of. There are also concerns about whether one can sustain holding that these two factors constitute a single state (Phillips, 2014).

Because higher-order states are rarely themselves conscious, we know about term as theoretical entities posited to explain how it is that some mental states are conscious. So, the posits made by different higher-order theories rarely themselves figure in subjective awareness. These higher-order theories have the status of scientific theories. They are not what are sometimes called folk theories, which are simply theoretically organized aspects of common sense.

So a disposition to have a higher-order thought and an actual higher-order thought that is not conscious will be subjectively indistinguishable. One cannot assess the merits of the various higher-order theories by differences in how the higher-order states occur subjectively. One must assess their merits by relevant theoretical implications.

Because most higher-order theories posit a higher-order awareness that is distinct from the first-order state it makes one aware of, some have objected that the higher-order state could misrepresent the mental properties of the first-order state and even that a higher-order state could make one aware of being in a first-order state that does not occur at all (Block, 2011; Levine, 2001). But higher-order theories do not predict or imply that those things do ever occur. So a higher-order theory could simply add a stipulation that they never do, so that misrepresentation by subjective awareness would no longer be possible on the theory. Because such an easy, obvious fix is available, it is likely that the real concern objectors have is that a theory of consciousness should rule that such occurrences are impossible on their own, without any such stipulation. But those who advance this objection do not make it clear why this should be so.

In any case, the objection from misrepresentation is arguably misguided. For one thing, it arguably rests on a taking the consciousness of a mental state to be intrinsic to that state, which might make it seem that misrepresentation by subjective awareness could not occur (Weisberg, 2011). But conceiving of a mental state's being conscious in that way is at best controversial and also conflicts with most higher-order theories.

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 higher-order misrepresentation.

The objection from misrepresentation also ignores an important advantage of any higher-order theory. Consciousness is a matter of mental appearance. It is the way our mental lives subjectively appear to us.

The higher-order states posited by higher-order theories 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). Higher-order theories dispute that since a higher-order state would constitute the reality of consciousness, and the content of that higher-order state would constitute the subjective appearances. In any case, consciousness does not exhaust mental functioning. So, mental appearances of consciousness are distinct from the underlying mental reality that occurs independently of consciousness (Rosenthal, 2022).

Broader connections

Higher-order theories always posit two factors: a higher-order factor that explains the subjective appearances and a first-order factor that explains psychological functioning independently of those appearances. So higher-order theories are especially useful in explaining occurrences in which subjective awareness diverges from psychological functioning. One example 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. Higher-order theories readily explain such occurrences. It is unclear how any other theory of consciousness could explain actual occurrences in which subjective awareness does diverge from psychological functioning.

There are other examples in which that does happen. In blindsight, damage to the primary visual cortex prevents subjective awareness of stimuli even though those stimuli demonstrably affect psychological functioning. Again, higher-order theories readily explain how this can occur (Weiskrantz, 2009). There are other neurological and psychological considerations that also provide strong support for a higher-order theory. For example, there is the finding of a neural difference specifically in the prefrontal cortex between conscious and unconscious cases since the higher-order awareness posited by higher-order theories almost certainly would occur in the prefrontal cortex (Lau & Passingham, 2006; see also Lau & Rosenthal, 2011).

Further reading

- Gennaro, R. J. (Ed.). (2004). *Higher-order theories of consciousness*. John Benjamins Publishers.
- Weisberg, J. (2020). Higher-order theories of consciousness. In U. Kriegel (Ed.), *Oxford handbook of the philosophy of consciousness* (pp. 438–457). Oxford University Press.
- Weisberg, J., & Rosenthal, D. (in press). Higher-order theories of consciousness. In A. Raffone (Ed.), *Consciousness: A comprehensive reference* (2nd ed.). Elsevier.

References

- Armstrong, D. M. (1993). *A materialist theory of the mind* (2nd ed.). Routledge. (Original work published 1968)

  ↩

- Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*(5–6), 481–499. https://doi.org/10.1017/S0140525X07002786

  ↩

- Block, N. (2011). The higher order approach to consciousness is defunct. *Analysis, 71*(3), 419–431. https://doi.org/10.1093/analys/anr037

  ↩

- Carruthers, P. (1996). *Language, thought and consciousness: An essay in philosophical psychology*. Cambridge University Press.

  ↩

- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79*(1–2), 1–37. https://doi.org/10.1016/S0010-0277(00)00123-2

  ↩

- Fernandez-Duque, D., & Thornton, I. M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition, 7*(1–3), 324–344. https://doi.org/10.1080/135062800394838

  ↩

- Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception* (pp. 89–110). Oxford University Press.

  ↩

- Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. Oxford University Press.

  ↩

- Lau, H. (2022). *In consciousness we trust*. Oxford University Press.

  ↩

- Lau, H., & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences, 103*(9), 18763–18768. https://doi.org/10.1073/pnas.0607716103

  ↩

- Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009

↩

- Leibniz, G. W. (2014). *Monadology* (L. Strickland, Trans.). Edinburgh University Press. (Original work published 1686)

↩

- Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford University Press.

↩

- Locke, J. (1975). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press. (Original work published 1689)

↩

- Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1755). https://doi.org/10.1098/rstb.2017.0357

↩

- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review, 83*(4), 435–450. https://doi.org/10.2307/2183914

↩

- Nietzsche, F. (2001). *The gay science* (B. Williams, Ed.; J. Nauckhoff & A. Del Caro, Trans.). Cambridge University Press. (Original work published 1882)

↩

- Phillips, B. (2014). Indirect representation and the self-representational theory of consciousness. *Philosophical Studies, 167*(2), 273–290. https://doi.org/10.1007/s11098-012-0087-1

↩

- Proust, J. (2013). *The philosophy of metacognition: Mental agency and self-awareness*. Oxford University Press.

↩

- Riccardi, M. (2021). *Nietzsche's philosophical psychology*. Oxford University Press.

↩

- Rosenthal, D. (1986). Two concepts of consciousness. *Philosophical Studies, 49*(3), 329–359. https://doi.org/10.1007/bf00355521

↩

- Rosenthal, D. (1997). A theory of consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates* (pp. 729–753). MIT Press.

  ↩

- Rosenthal, D. (2005). *Consciousness and mind*. Clarendon Press.

  ↩

- Rosenthal, D. (2022). Mental appearance and mental reality. In J. Weisberg (Ed.), *Qualitative consciousness: Themes from the philosophy of David Rosenthal* (pp. 243–271). Cambridge University Press.

  ↩

- Weisberg, J. (2011). Misrepresenting consciousness. *Philosophical Studies, 154*(3), 409–433. https://doi.org/10.1007/s11098-010-9567-3

  ↩

- Weiskrantz, L. (2009). *Blindsight: A case study spanning 35 years and new developments* (2nd ed.). Oxford University Press.

  ↩