CellPress REVIEWS

Opinion

Understanding the Higher-Order Approach to Consciousness

Richard Brown,[1] Hakwan Lau,[2,3] and Joseph E. LeDoux[4,5,6,*]

The higher-order theory (HOT) of consciousness has often been misunderstood by critics. Here, we clarify its position on several issues, and distinguish it from other views, such as the global workspace theory (GWT) and early sensory models (e.g., first-order local recurrency theories). For example, HOT has been criticized for overintellectualizing consciousness. We show that, while higher-order states are cognitively assembled, the requirements are in fact considerably less than often presumed. In this sense, HOT may be viewed as an intermediate position between GWT and early sensory views. We also clarify that most proponents of HOT do not stipulate consciousness as equivalent to metacognition or confidence. Furthermore, compared with other existing theories, HOT can arguably account better for complex everyday experiences, such as emotions and episodic memories. This makes HOT particularly useful as a framework for conceptualizing pathological mental states.

Highlights

Misunderstandings about HOT have marginalized it relative to other approaches.

We clarify some of the key misunderstandings, including assumptions about the equivalence between consciousness and metacognition, and the role of introspection and the self.

We reply to several objections, including those concerning the nature of perception in the visual periphery and issues raised about the value of so-called 'no report' paradigms.

We also address issues regarding the involvement of prefrontal cortex, including questions about whether it is necessary, and whether its deactivation during dreams and psychedelic states is incompatible with it contributing to higher-order awareness.

We propose a reconceptualization of lower-order states that contribute to higher-order awareness, including states of prefrontal cortex and multimodal and mnemonic states processed in posterior cortical areas.

We provide arguments as to why the HOT of consciousness may be superior to both GWT and local recurrency theory regarding its ability to account for subjective experiences, especially of complex states such as memories and emotions that occur in everyday life and that are hallmarks of psychopathological conditions.

What Is HOT?

Consciousness (see Glossary), as used here, refers to subjective experience, or what is sometimes called **phenomenal consciousness** [1,2], as opposed to the condition of merely being awake and alert and behaviorally responsive to external stimuli. To be phenomenally conscious is for there to be something that it is like to be the entity in question, that is, something that it is like for the entity itself [2].

Subjective experience is the stuff of novels, poems, and songs, of our emotions and memories, the essence of being a human. It is hard to imagine what it would be like to not be sentient in the way we are. Unsurprisingly, then, the science of consciousness is currently a vibrant and thriving area of research. However, there is no generally accepted theory of the phenomena being studied, and the phenomena themselves often do not include many of the kinds of complex experiences that we normally have in the course of day-to-day life, such as of our emotions and memories.

Here, we argue that the foundation for a viable theory of such experiences exists but has not been given the credit it deserves. We are referring to various ideas known collectively as HOT. HOT is often overlooked, or lumped in with other cognitive theories of consciousness, such as GWT. Even worse, it is sometimes dismissed out of hand as not empirically plausible compared with other theories. Our main goal here is to identify and answer some of the most common questions concerning the theory. First, however, we briefly describe key features of HOT.

Theories of consciousness can be categorized as either first-order theories or HOTs. In general, there are two features that make a theory of consciousness a HOT rather than a first-order theory. The first feature is a commitment to the claim that a mere **first-order representation** is not sufficient for conscious experiences to arise; some higher-order mechanism or mechanisms

[1]Philosophy Program LaGuardia Community College, City University of New York, Long Island City, NY, USA
[2]Department of Psychology and Brain Research Institute, University of California, Los Angeles, CA, USA

 https://doi.org/10.1016/j.tics.2019.06.009

are also needed. For example, having a first-order perceptual state about an external object is often crucial for the organism to respond meaningfully to it. However, according to HOT, such first-order states can occur nonconsciously and are insufficient for phenomenally conscious experiences to occur [3]. This requirement of HOT distinguishes it from local versions of first-order theories, such as recurrent processing theories [4–7].

The second requirement of HOT comes from the common-sense claim that if an organism is in no way aware of itself as being in a certain first-order state, then it is not phenomenally conscious of the content of that state. This is a logical consequence of what is called the **Transitivity Principle** [3]. The basic idea, according to HOT, is that conscious experiences entail some kind of minimal **inner awareness** of one's ongoing mental functioning, and this is due to the first-order state being in some ways monitored or meta-represented by a relevant **higher-order representation**. This requirement of HOT distinguishes it from cognitive theories, such as GWT [8–11], which also invoke additional cognitive processes as a crucial element of conscious experience, but which do not posit this type of inner awareness. For this reason, GWT is a variant of first-order theory.

HOT is not a single entity. It in fact comes in many varieties [12–14] that are distinguished by the different ways in which they define the relevant higher-order mechanism (Table 1). For example, the traditional view, sometimes attributed to John Locke and Immanuel Kant, refers to the mechanism of inner awareness as an Inner Sense, akin to perception [15,16]. However, Inner Sense HOTs, also referred to as Higher-Order Perception theories, have somewhat fallen out of favor because of a failure to find a neural implementation of an Inner Sense [17]. Another variety, Dispositional HOT, posits that the mere availability of first-order content to higher-order mechanisms accounts for consciousness [18]. Other versions of HOT include the Phenomenal Self theory [19], the Radical Plasticity Hypothesis [20], and variants of Higher-Order Thought Theory (HOTT) [3,12,14,21–24,49]. HOTT postulates that the higher-order state is thought-like, and is the subject of much current discussion and debate [5,23,25–30].

Current Status of HOT

HOT emerged from a philosophical theory, but is not simply an intriguing philosophical hypothesis. It is also broadly compatible with empirical findings in neuroscience. For example, under highly controlled conditions, a difference in perceptual awareness is better correlated with activity in brain areas responsible for high-level cognition rather than in early sensory regions [31,32]. Also, disruption of such activity by magnetic stimulation [33,34] or lesion [35] can change subjective aspects of the perceptual experience. These findings have been reviewed previously [13,36], and there have also been recent updates [23,37], in which some standard philosophical objections have also been addressed.

Here, we discuss some of these findings and replies to objections in details. However, instead of providing yet another broad review, we go through these issues in the context of clarifying some common misunderstandings by framing them in terms of questions about HOT. Given that much of the evidence supporting HOT is also compatible with other cognitive theories of consciousness, such as GWT, it has not been obvious how to arbitrate between the two. Our goal is to make clear this distinction to allow further experiments to arbitrate between the different theories.

Does HOT Require Sophisticated Thoughts?

According to HOT, cognitive access, in the form of a kind of inner awareness, is necessary for phenomenal consciousness. A common objection then is that HOT makes consciousness require overly sophisticated mechanisms [4,38].

[3]Department of Psychology, and State Key Laboratory for Brain and Cognitive Sciences, the University of Hong Kong, Hong Kong, China
[4]Center for Neural Science and Department of Psychology, New York University, New York, NY, USA
[5]Departments of Psychiatry and Child and Adolescent Psychiatry, New York University Langone Medical School, New York, NY, USA
[6]Nathan Kline Institute, Orangeburg, NY, USA

*Correspondence:
jel1@nyu.edu (J.E. LeDoux).

One version of this criticism is that HOTT appears to commit itself to the cognitive requirement of having concepts about mental states that constitute the content of higher-order thoughts. Various HOTTs respond differently to this challenge. Some appeal to innate concepts. Others deny that the relevant conceptual capacities are sophisticated [3]. Still others question whether concepts are required at all [23].

Another version of this objection argues that, given the richness of everyday experience, and the fact that the human capacity for thoughts is limited, consciousness under HOTT would require too many higher-order thoughts [18]. However, the exact level and degree of complex processing required to generate relevant higher-order thoughts is not currently known. Some higher-order theorists propose that the relevant higher-order states may use relational concepts ('*x* is bluer than *y*') [3], which would reduce both the amount of processing and number of thoughts required relative to what might at first be expected.

Confusion also results in part from the ways 'thought,' 'introspection', and 'self,' as used in HOT, are interpreted by critics [28]. Traditionally, **introspection** refers to an active process in which one becomes conscious of their inner states [39,40]. Similarly, 'self' is often used to refer to conscious awareness of one's self. The fact that higher-order philosophers tend to use propositional statements involving personal pronouns ('I see red') to describe the higher-order thought has led to the idea that HOT implies a conscious self that has introspective knowledge of its experiences [41]. However, HOTT proponents, in fact, typically call upon a cognitively 'lean' conception of both thought and self. In some versions, a monitoring mechanism that works at the subpersonal level is postulated [42], while, in other versions, the thoughts are at the personal level but are arrived at automatically and without appearing to be the product of inference [3]. Importantly, both the leaner/passive and thicker/active versions of self and introspection relevant to HOT are distinct from more basic biological mechanisms referred to with terms such as 'self as object' (as opposed to subject) [43,44], 'core self' [45], or the 'machinery of self' [46]. As such, the thoughts involved in higher-order representations are not usually themselves viewed as ones we are conscious of being in [3,15,23,24,36,47–49]. A further step, involving a more elaborate, active, form of introspection is typically required to be conscious of the higher-order state.

While HOTs are typically agnostic regarding the status of consciousness in animals, the leaner sense of introspection and self makes it an open question as to whether animals are conscious under HOT. Many have argued for unique aspects of human cognition [50–52]. We suggest that Tulving's distinction between autonoetic (self) awareness and noetic (sematic) awareness [24,48,49,53,54] offers a way of conceptualizing what kinds of phenomenally conscious experience might be possible via cognitive re-representation in humans versus other animals. For example, while only humans may experience autonoetic (self) consciousness (awareness of one's self as the subject and owner of the experience) [48,49,53], some other animals, especially non-human primates [55–57], may, similar to humans, have higher-order noetic conscious states (e.g., awareness of perceptual, conceptual, or memory representations) [49]. Still other animals may have more primitive forms of awareness (e.g., of body states) that do not involve higher-order representation [45,58]. Nevertheless, demonstrating conscious awareness in animals is methodologically challenging [59–62] since most studies rely on mere analogy with human behavior [45,63–67]. Similar issues and complications apply to questions about scientifically determining whether young infants or robots may be conscious and, if so, in which way (or ways).

Does HOT Equate Consciousness with Metacognition or Confidence?

In psychological experiments on **metacognition**, subjects are often required to give confidence ratings after a simple perceptual or memory task, indicating how sure they are that their decisions are

Glossary

Blindsight: the neurological phenomenon that, after lesion to primary visual cortex, patients deny having conscious experience for the affected regions in the visual field, and yet they can detect and discriminate visual targets presented there above chance levels.
Consciousness: term used to refer to different mechanisms and phenomena, such as whether a person is awake versus in a coma, or is having subjective experiences. Here, we are primarily concerned with states that have subjective phenomenal qualities, such as perceptions, memories, thoughts, or emotions.
First-order representations: generally **mental representations** about states of the world, or about one's situation in the world. In the case of visual perception, for example, these involve activity in striate and extrastriate cortical areas, where neuronal firing can be directly driven by external stimuli, such as lines, color, motion, and objects. First-order theories of consciousness include early sensory/local recurrency and global broadcasting theories, and do not require any additional higher-order representations.
Higher-order representations: generally about other mental representations, such as first-order representations, and are usually associated with areas of PFC. Higher-order theory typically proposes that the higher-order representation, while necessary for consciousness, is itself not one of which you are aware. To be aware of a higher-order representation requires an additional higher-order representation.
Inner awareness: occurs when one is aware of one's own mental representations, a process that we propose involves higher-order representations.
Introspection: in typical usage, this refers to content resulting from the act of intentional and effortful self-monitoring and evaluation of ongoing conscious experiences, as when one directs one's conscious thoughts to other conscious states. HOT does not require this type of elaborate, active, conscious introspection for simple experiences. Instead, they appeal to a relatively passive (cognitively 'leaner') process of metarepresentation, which operates at a nonconscious level.

correct. While these tasks are useful experimental devices for probing mechanisms relevant to conscious awareness, most proponents of HOT do not in fact treat metacognition and consciousness as conceptually equivalent [68,69]. Metacognition, in this sense, is typically a process we are conscious of ourselves as engaging in, whereas the higher-order processes postulated by HOT to underlie inner awareness are typically ones we are not conscious of ourselves as engaging in.

Why then do proponents of HOT use explicit metacognitive tasks in experiments? One reason may be related to an empirical observation. Evidence suggests that neural activity in prefrontal cortex (PFC) is important for conscious perception [70]. When such activity is manipulated with transcranial magnetic stimulation (TMS), explicit (conscious) metacognition is also affected [33,34,71]. This suggests that the brain mechanisms responsible for higher-order conscious experience and explicit metacognition are shared. Perhaps metacognition appeared first in evolution as a process unrelated to consciousness, and then the same mechanism was later 'recycled' to enable consciousness. In this sense, explicit metacognitive tasks are a convenient tool for studying mechanisms relevant to consciousness, even if these tasks do not necessarily represent the *sine qua non* of consciousness.

For other proponents of HOT, there may be deeper conceptual reasons why some specific metacognitive questions conveniently tap into the relevant higher-order states. For example, it has been argued that, when a subject says, 'I see a red square', they are reporting a higher-order thought; that is, they are reporting that they have a thought to the effect that 'I see a red square'. Some have argued [3] that this allows us to explain why we can verbally report our conscious states, but cannot verbally report unconscious mental states. Thus, to be able to report an unconscious state, one would have to have the relevant higher-order thought. However, that would mean that the state was conscious. So, in some cases where one expresses confidence in a judgment about seeing a red square, arguably the confidence also expresses a higher-order thought. However, while confidence is generally a good guide, many higher-order theorists still consider verbal report of awareness, not confidence, to be the gold standard for showing that one is in a conscious state [72].

Relatedly, some claim that metacognitive tasks can be performed with nonconscious (subliminal) stimulus presentations (Box 1). This may be perceived as a challenge to HOT because it might be viewed as an empirical example where there is higher-order awareness and/or confidence but no conscious experience. However, HOT itself is agnostic as to whether this form of metacognition can occur. Since not every metacognitive judgment is of the kind postulated to engender conscious experience, it is an open question as to whether there may be some kind of metacognition directed at unconscious states.

What Is the Difference between HOT and GWT?

Similar to HOT, variants of GWT also suggest that early sensory representations alone are insufficient for consciousness; additional processes are needed in both of the theories. For HOT, such processes involve higher-order representations instantiating a kind of inner awareness, whereas, for GWT, the essential processes boost and stabilize sensory signals. Thus, GWT is strictly a first-order theory; the broadcast signals are not necessarily metarepresentations (i.e., they are propagations of essentially the same signal rather than a new representation about the relevant first-order representation). Therefore, while the two theories are somewhat related, they differ in important ways.

In a recent review, global broadcasting and higher-order monitoring were cast as two orthogonal dimensions of processing [73]. The authors agreed that both kinds of process, especially their conjunction, are typically involved in everyday conscious experiences; if one were to speculate

Mental representation: can be conscious or nonconscious, and includes, for example, sensory, perceptual, mnemonic, or conceptual information.

Metacognition: formally, this refers to a cognitive process about another cognitive process. A task can tap into a suprapersonal form of metacognition by asking subjects to provide explicit judgments about some aspect of another process (e.g., confidence about how well one can discriminate a visual target). However, metacognition can also happen implicitly, at a subpersonal level (e.g., some metacognitive process is involved when we spontaneously notice that we have made an error without being required to do so, or when we effortlessly recognize that a certain visual image is caused by our own imagination rather than triggered by an external input).

Phenomenal consciousness: when we say that an individual is phenomenally conscious, we mean, following Thomas Nagel, that they are capable of having states for which there is something that it is like, from the individual's point of view, to be in such states. Although the term is sometimes equated with first-order states, it is, in fact, itself neutral with respect to whether phenomenal consciousness depends on first- or higher-order representations. Our theoretical position is that the higher-order account is correct.

Transitivity principle: assumes that when you are in no way aware of being in some mental state, there is nothing that it is like for you to be in that mental state. For example, if you are presented with a red stimulus but are in no way aware that you are seeing red, you are not consciously experiencing red. The converse of this is the transitivity principle with which HOT is concerned, that consciously seeing red depends in some way on one being aware of having that experience.

Table 1. Differences between HOTs and Other Theories of Consciousness

Theory[a]	Is it a variety of HOT?	Is first-order content re-represented?	Are complex 'thoughts' required?	Is conceptual processing involved?	Is explicit metacognition required?	What is the function of consciousness in the theory?
Higher-Order Thought Theory (HOTT) [3,13]	Yes	Yes, but redescribed in different format	Yes	Yes	No	Little is assumed by the theory
Higher-Order Representation of a Representation (HOROR) [22–24]	Yes	Yes, possibly redescribed in different format	Not clear	Yes	No	Whatever the function of the right kind of higher-order representations of representations turns out to be
Perceptual Reality Monitoring Theory [42]	Yes	No	No, relevant higher-order states are mere indexes	No	No	Formation of subjectively justified beliefs
Multi-State Hierarchical Model of Subjective Self-Awareness [49]	Yes	Diverse hierarchically organized lower-order states are re-represented	Not required, but can result	Yes, in the form of schema	No	Thoughts, beliefs, memories, feelings, and attributions about the world, and about oneself as an object and a subject
Radical Plasticity Thesis [20]	Yes	Yes, but redescribed in different format	No, relevant higher-order states are not thoughts	Typically no	No	Control of behavior and learning in novel situations
First-order local recurrency view [1,6,78,87]	No	No	No	No	No	Minimal
Global Workspace Theory (GWT) [8–11]	No	Propagated but not necessarily metarepresented	No	No	No	Global sharing and stabilization of information
Global Workspace Plus Metacognition [50,131]	No	No	No	No	Yes	Global sharing and stabilization of information and interpersonal communication

[a]This table only covers a few theories to highlight contrasts between different HOTs and between HOT and other theories in the literature. The issues in question are as follows. (i) Is the theory a variety of HOT? (ii) Does first-order content need to be re-represented at a later stage to become conscious? (iii) Does everyday experience mean that many thought-like representations are involved at one moment? (iv) Does consciousness require capacity for conceptual processing? (v) Does consciousness require explicit (contra subpersonal-level) metacognition? (vi) What is the function of consciousness?

about the way in which conscious artificial intelligence could be built, both dimensions would be relevant. However, while global broadcasting tends to go hand-in-hand with higher-order awareness in typical cases, this is not necessarily true for all conscious experiences. For example, in peripheral vision, one has the sense of consciously seeing the details, but the details are typically not globally accessible as such. HOT accounts for this by interpreting such cases in terms of higher-order inflation (Box 2) and, thus, without requiring actual broadcast of the detailed content.

The contrary may also occur: information that is globally broadcasted may not be subjectively experienced. For example, imagine maintaining images of a few abstract shapes to respond a few seconds later about what the shapes are. During the delay, active vivid imagery may not be necessary. In fact, vividness of imagery varies between individuals [74], but even those who do not experience anything like normal seeing during the delay (i.e., they are not vivid imagers)

Box 1. Unconscious Working Memory and Metacognition

The possibility that working memory and metacognition may operate nonconsciously is sometimes taken as a challenge to HOT. In the main text, we argued that these claims, even if true, pose no threat to HOT *per se.*

Independently, it is also important to point out that claims of explicit metacognition with nonconscious stimuli (as opposed to implicit or unconscious metacognition) are still not totally convincing; these hinge on how unconsciousness is defined.

As has been pointed out by decades of psychophysics, simply asking people to label trials within a task condition as 'consciously seen' versus 'unconscious' is not always sufficient [132]. There is a serious concern about an arbitrary criterion being set to fit trials into the two forced labels. At-chance discrimination is traditionally considered a conservative criterion for showing that the stimuli are not consciously perceived [132] and, as blindsight shows, above-chance discrimination does not mean that the relevant stimuli are consciously perceived. Alternatively, one can try to show that, although such stimuli can be discriminated better than chance, they are subjectively indistinguishable from a 'blank' stimulus containing no information [133]. However, current efforts to demonstrate 'unconscious' working memory [134] and unconscious metacognition [133] have not passed either of these two criteria for demonstrating unawareness.

Overall, to the extent that they are convincing demonstrations, studies of unconscious processing are constrained by the fact that such stimuli are typically weak (to pass either of the above criteria) [135]. Thus, as with the discussion of animal consciousness (see main text), it is important to keep in mind that there may be a gap between what we can measure scientifically and what is indeed conscious or not. That is, the fact that it is difficult to demonstrate nonconscious working memory and metacognition does not rule out the possibility that they exist.

tend to have no problem holding the information [75]. Given that the information is globally accessible in both cases (seeing versus holding information online in memory with and/or without strong imagery), what accounts for the difference in phenomenology? This is unlikely to be a simple matter of differences in richness of information between top-down versus bottom-up processes. For example, in dreams and hallucinations, when there is also no bottom-up input, the subjective experience is still vivid. HOT suggests this is due to the higher-order processes. While these cases do not decisively prove that GWT is wrong, they do raise questions regarding whether global availability and informational richness along can account for the difference in phenomenology. These cases also suggest possible experimental way of arbitrating between the theories.

Box 2. Subjective Inflation in the Unattended Periphery

One's impression of experiencing details is not necessarily fully veridical. We have the subjective impression of seeing vivid color in the periphery despite our relatively poor early sensory processing. According to proponents of HOT, our subjective impression may be at times inflated beyond the actual representational capacity [136], due to misrepresentation at the higher-order level. This is different from the more traditional account of 'filling in' [137], which suggests that actual details are added at the first-order sensory level. In the higher-order account, the first-order details may remain scant. The two mechanisms need not be mutually exclusive. The subjective richness may be due in part to first-order filling-in, augmented by higher-order inflation.

Against this view, Ned Block [5] argued that color perception is in fact not weak in the periphery if we enlarge the stimuli for the periphery to compensate for the larger receptive field size there [23]. However, we are not sure how this point is relevant because, in real life rather than in calibrated psychophysics experiments, stimuli do not automatically enlarge themselves upon entering our periphery. What matters is that, for the same stimuli, color processing is relatively weak in the periphery, and yet we see the world as somewhat rather uniform.

Besides color [138], inflation for other attributes are also well documented: for example, subjects reported higher false alarm rates for detection in the periphery, or in unattended locations. Critics argue that these results only reflected a bias in responding strategies, and it would beg the question to assume that such biases necessarily reflect inflated subjective experiences [5]. However, it is unclear whether the charge of begging the question is fair, because proponents of HOT do provide independent reasons for believing that the inflation results reflect experiences rather than just responses. For instance, if inflation reflects a biased strategy, it would be difficult to explain why the results persisted even upon training with extensive feedback. The hypothesis that subjective experience itself inflates appears to account for the data as well as anecdotal experience better.

Is HOT Merely a Theory of Access Consciousness?

Some proponents of GWT posit that global broadcast is a mechanism for conscious access, implying that access rather than phenomenology *per se* is all one can experimentally study [76] (see [11] for a possible exception). However, while it is possible to construe HOT as merely a theory of access consciousness [27], many proponents of HOT argue that it is primarily about subjective experiences rather than cognitive access to information.

For example, experiments in which one can report phenomenally experiencing a large array of stimuli but is unable to report the details are often used to challenge HOT. It is argued that, because the mechanism that allows access is occupied with doing other tasks (i.e., overflowed with other content) [77], one is not able to access all of one's phenomenally conscious experiences at any given moment. Such evidence has been taken to support first-order views [78]. However, HOT also readily provides an account of these findings, precisely because HOT is about experience not reports *per se*. That is, participants' impression of rich and unreportable experience may be inflated beyond what the first-order states in fact represent. Subjects may experience these states as being rich and full of detail when in fact their conscious experience lacks this richness. We discuss this so-called 'subjective inflation' in Box 2.

A related line of criticism of HOT is that much of the empirical evidence supporting it comes from tasks requiring subjects to report about the stimuli, and the act of reporting could be a confound [79]. Therefore, the critics suggest the use of so-called 'no report' paradigms, in which subjects are not asked to make reports about the relevant stimuli [80]. Advocacy of such paradigms, in our opinion, has unfortunately generated undue excitement. It is, of course, important to control for experimental confounds. Indeed, this 'task demand' (i.e., the need for subjects to report about the relevant stimuli) has long been controlled for and addressed in the literature [81–83]. It is also true that, when subjects were required to direct their attention away from the stimulus, some studies failed to find a difference in the degree to which higher-order activity tracked conscious perception using basic neuroimaging analysis [79]. However, it is a mistake to generalize from these null findings, because studies using more sensitive methods showed clear involvement of higher-order mechanisms, even when the subjects were not required to report about the stimuli [37,84,85].

Is Consciousness Better Correlated with First-Order Content?

Perhaps one reason why critics often think that HOT is about report and access is that, pretheoretically, conscious experiences appear to be associated with first-order perception [1,86]. After all, activity in the visual cortex is typically understood to reflect the content in the conscious experiences concerned, even in some versions of HOT [23]. For example, if V4 processes information about color, and MT processes information about motion, then one might expect that these are the areas that subserve conscious experience of these properties [87]. Therefore, anything beyond first-order content areas may just reflect post-perceptual processing that allows cognitive access and reporting.

However, while it is true that the various visual areas appear to code relevant content, it is unclear whether their link to consciousness is straightforward. It is known that subliminal or nonconscious stimulation also engages similar sensory circuits as consciously experienced stimuli [88]. Also, ongoing spontaneous activity in the very same neurons that underlie normal perception is typically nonconscious [89]. One possibility is that such activity has to reach a certain threshold [90], or that it has to achieve the right sort of dynamic profile [89], such as involving feedback to V1. However, patients with damage to V1 can still have visual experiences [23,91]. So, at least from the outset, an alternative appears to be just as plausible: that (extrastriate) visual activity has to be re-represented by higher-order processes to become conscious, as suggested by HOT.

This is related to another common criticism of HOT concerning possible mismatches between higher-order and first-order contents [27]. This is only a concern in models in which the higher-order representation redescribes the content from the first level, which is not true of all HOTs [23]. These criticisms have also been discussed and addressed elsewhere [22–24]. Even so, these considerations suggest that the notion of a mismatch is not merely a philosophical objection to HOT but instead can be seen as an empirical predication of HOT. That is, HOT predicts that we should find cases where subjective experience varies (along with higher-order representations) while first-order content remains constant, for which there is some supporting evidence [23].

Does HOT Imply That Consciousness Lacks a Function?

While some proponents of HOTT argue that possible functions of consciousness are likely minimal [3], others do not necessarily agree [92]. Regardless, this does not mean that higher-order representations are epiphenomena; it simply means that HOT itself, as a theory, is not committed to making strong assumptions about what the functions of consciousness may be, without further empirical evidence. Thus, consciousness itself does not necessarily guarantee superior behavioral performance in a task, but it might in some situations [49,92].

This last point may be particularly relevant for understanding powerful forms of nonconscious processing, such as **blindsight**, in which patients deny having subjective experiences following damage to the visual cortex, even though they can guess the identity of visual stimuli well above chance level [93]. For GWT, this lack of conscious experience must mean that the relevant signal is not globally available to all major modules of the brain; somehow, some local pathway must have made possible the successful guessing and stimulus identification. Although we do not argue against the existence of such local pathway [94], it does raise the questions, if a local pathway can support behavioral performance too, what exactly is the benefit of global broadcast? Are these local pathways always unconscious?

According to HOT, the answer is no. A mere difference in the higher-order state determines whether a nonconscious perceptual process is consciously experienced [22,24,95]; that is, whether an early sensory signal is conscious does not depend on its anatomical locus alone (global or local); the very same sensory signal can be conscious or not depending on a higher-order process. This is why some nonconscious processes can be so powerful: they have all the functions of the first-order components of a conscious process. Yet, this does not necessarily mean that the additional higher-order process makes no additional contribution. However, what this contribution might be depends on the variant of HOT.

Does HOT Propose That Consciousness Is in the Prefrontal Cortex?

HOT is not about PFC *per se.* It is a theory about fundamental cognitive requirements of consciousness. Recent advances in cognitive neuroscience have clearly implicated PFC in higher cognitive mechanisms [96–98], such as metacognitive mechanisms that underlie higher-order representations. These findings are consistent with our hypothesis that PFC is a key part of the network with a crucial role in generating the relevant kind of higher-order cognitive states that underlie phenomenal awareness, as depicted in Figure 1 and described further in Box 3.

However, as in the discussion of the role of sophisticated thoughts in HOT, different HOTs do not agree on what PFC contributes to consciousness. Some proponents of HOTT have suggested that a conscious first-order visual state involves one having a certain specific kind of PFC activity

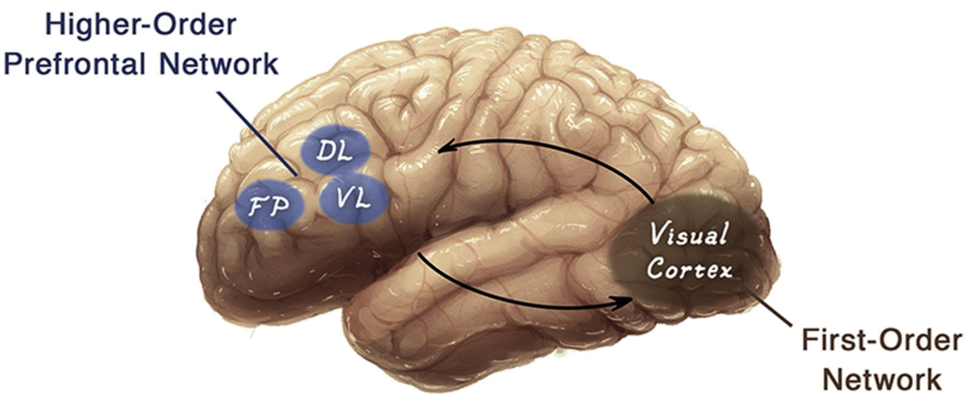

Figure 1. Prefrontal Network Proposed to Underlie Higher-Order Perceptual Awareness. In higher-order theory (HOT), neural states of First-Order networks are viewed as nonconscious representations that are rendered conscious when re-represented by the higher-order network involving areas of prefrontal cortex. This model mainly applies to perceptual awareness. See Figure I in Box 3 for ways in which this model might be modified to account for higher-order awareness of conceptual, mnemonic, and emotional information. Abbreviations: DL, dorsal lateral; FP, frontal pole; VL, ventral lateral.

[36]. A related view is that higher-order prefrontal representations index the relevant first-order content in sensory areas to allow that content to enter consciousness [23]. In this sense, consciousness is 'jointly determined' by concurrent activity in the higher-order areas and areas that maintain the lower-order representations. Alternatively, in higher-order representation of a representation (HOROR) theory, the relevant higher-order state in PFC itself is phenomenally conscious [22,24,95]. The multistate hierarchical model proposes that diverse kinds of lower-order states are represented and integrated with self-information to give rise to higher-order conscious states involving one's self [49]. Which of these, if any, might be correct is an open question (Table 1). Despite these differences between theories, most variants of HOT are united in their agreement that PFC has a key role in consciousness; but proponents of HOT do not typically assume that one is 'more conscious' whenever there is more PFC activity. Unlike early sensory neural representations, PFC neurons do not straightforwardly signal constant stimulus features; neuronal coding may be complex and involve local and distributed ensembles within and between areas of PFC [99] (also see Figure I in Box 3). As such, observations that activity in a given PFC region, as measured with neuroimaging, may decrease while one is dreaming [100], or in a psychedelic state [101], do not pose a challenge to HOT; such a simple relationship at this coarse-grained level is not generally presumed to be the case by HOT.

In evaluating current data, it is important to note that 'PFC' is a generic conception that subsumes several different subregions and functional specializations [49,102–108]. Unfortunately, this diversity has not always been kept in mind when discussing the involvement of the PFC in HOT or other theories about consciousness. Some have even argued against constitutive involvement of PFC in any form of consciousness, using findings from cases with partial lesions of some areas of PFC [109]. However, the anatomical details reported were at times inaccurate [37]. Box 3 suggests how the heterogeneity of PFC, and the rich diversity of its inputs, can be leveraged to achieve a broader, more nuanced view of higher-order representations in consciousness.

Why Is HOT Needed?

Much of the debate about the nature of consciousness and its neural underpinnings has concerned visual perception. While we believe that HOT has advantages in this area, we propose that it is especially useful in accounting for everyday experiences (e.g., appreciation of art, music, or poetry; reminiscing about our past; experiencing emotions; or reflecting upon our self).

Emotion, which has largely been ignored in the modern scientific study of consciousness [24], is a case in point.

Recent theories and research implicate cognition and conceptual knowledge as being central to conscious emotional experience [48,49,110–113]. Cognitive theories of emotion are also important in understanding psychopathological conditions, which often crucially involve thoughts about one's life problems (e.g., [114–116]. Therefore, HOT, and its cognitive underpinnings, may offer a framework for understanding both healthy and maladaptive emotions. The HOROR variant of HOTT [22,23] has, in fact been used to account for emotional consciousness [24].

For decades, the subjective experiences of patients with affective disorders have been marginalized when evaluating the effectiveness of behavioral, cognitive, and pharmaceutical treatments [117]. In medical models, for example, the goal has been to find biological correlates that, if changed, especially by drugs, will solve the problem [48,118]. However, these efforts to develop novel agents through studies of animals have been so unsuccessful that the pharmaceutical industry has significantly scaled back research in this area [119].

Such considerations underlie recent debates about whether targeting subcortical circuits, such as those of the amygdala, will help to alleviate symptoms of fear and anxiety in disorders such as phobias [118,120,121]. Of course, in a sense, the amygdala is important and relevant for fear disorders. However, If, in both mental health and psychoample evidence also suggests that it is unlikely to be the full story [24,48,117]. It is known that threats can elicit amygdala activity and trigger physiological responses nonconsciously [122,123]. Participants do not typically spontaneously report feelings, but even when asked for a verbal report, they do not respond in

Box 3. Prefrontal Cortex and Higher-Order Experience

'PFC' subsumes several different subregions (some located laterally and some medially) with different cytoarchitectonic properties (granular, dysgranular, or agranular), different patterns of connectivity with each other and with sensory, memory, and conceptual processing regions, and different phylogenetic histories, with some present in all mammals, some mainly in primates (dorsolateral), and some particularly well developed or perhaps even unique in humans (aspects of the frontal pole) [102–107]. This complexity, although at first sight appears daunting, might in fact offer opportunities for a more nuanced view of how PFC contributes to experience.

Lau and Rosenthal [13] presented a proposal for the role of specific areas of PFC in higher-order representations. Key areas in their model were the dorsolateral and polar regions of PFC. While their review was primarily about visual perception, adding the ventral lateral PFC allowed their model to generalize to other external senses [39] (Figure 1). Consistent with Rosenthal's HOTT [3], they suggested that higher-order representations in this network are not themselves consciously experienced, but instead facilitate the experience of first-order states. In their model, conscious experience of the higher-order representation itself requires an additional level of higher-order representation (not shown); although Lau and Rosenthal did not identify the source of the additional higher-order representation, the highly conceptual nature of polar prefrontal region makes it a candidate worth considering.

For the higher-order network model to account for more complex kinds of experience (e.g., of memories, emotions, and self), it needs to be revised to include lower-order states related to such experiences (Figure Ia). Given that some of the areas involved are themselves prefrontal areas, the nature of lower-order states may need to be reconceived [49]. For example, medial and insula prefrontal areas receive inputs from conceptual, memory, and subcortical circuits, as well as from sensory/perceptual inputs, and, in turn, connect with lateral-polar prefrontal areas. In other words, these prefrontal areas may construct lower-order, representations used by the lateral-polar prefrontal areas in the assembly of higher-order representations that are phenomenally experienced. If so, despite the fact that the prefrontal areas in question are higher-order anatomically and cognitively, they are lower order in relation to the higher-order network.

Alternatively, it is possible that the traditional higher-order network has been conceived too narrowly, and should include medial and insula prefrontal areas (Figure Ib). Another possibility is that the higher-order network does not involve a fixed set of prefrontal areas but instead a coalition of areas and connections that is flexibly recruited on a situational basis to meet the needs of the moment [49], not unlike the working memory collation hypothesis [139]. Alternatively, the higher-order network might extend beyond PFC and also involve posterior cortical areas in the parietal and/or temporal lobes.

CellPress
REVIEWS

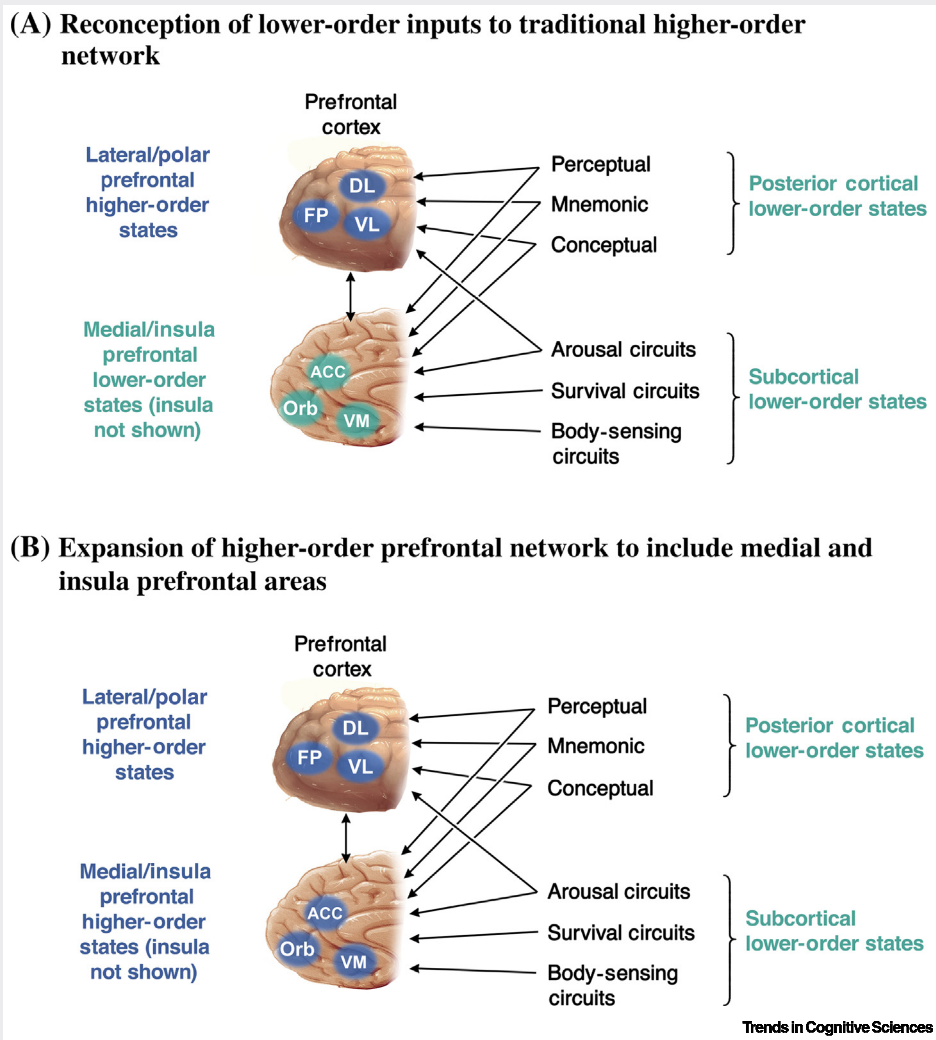

Figure I. Proposed Variants of the Prefrontal Higher-Order Network. (A) The prefrontal higher-order network is extended to account for conceptual, mnemonic, and emotional awareness. To do this, the nature of the lower-order states is reconceived, because some of the inputs to the traditional higher-order network are from other prefrontal areas, especially from medial and insula cortex (note that medial areas shown are only some of the possible relevant areas; also note that the insula region is buried below the cortical surface and is not depicted in these illustrations). (B) An alternative to the view in (A) is that all of the prefrontal areas indicated are part of the higher-order network, and that the lower-order inputs are all from nonprefrontal areas. As described in the main text, it is also possible that areas of the higher order prefrontal network are flexibly recruited on a situation-by-situation basis, and it is also possible that parietal and/or temporal cortical areas are involved. Color code: blue letters and shading depict higher-order, and green letters and shading lower-order, areas. Abbreviations: ACC, anterior cingulate cortex; DL, dorsal lateral; FP, frontal pole; Orb, Orbital; VL, ventral lateral; VM, ventromedial.

a way that would suggest they experienced an emotion appropriate to the eliciting stimulus [124]. Furthermore, direct electrical stimulation of the amygdala reliably elicits physiological responses, but only rarely subjective experiences [125]; even when subjective experiences are elicited, it is unclear that these arise from the amygdala itself, as opposed to resulting from activity spreading to higher-order processes [48]. Also, patients with amygdala lesions can consciously report emotional experiences [126], including fear [127]. If conscious emotions are higher-order cognitive cortical states, this might explain why antianxiety medications developed by assessing effects on behaviors controlled by subcortical circuits in animals have not been as successful as once hoped in changing human subjective experiences.

Thus, HOT provides a principled account to address what may be missing: to effectively treat problems related to fear and anxiety, the higher-order mechanisms of consciousness involving various areas of PFC, and the lower-order mechanisms involving subcortical defensive circuitry, may both have to be treated [48,118]. However, to do so effectively, each may have to be targeted separately, otherwise either one may reinstate the other [48,117,118,121].

In a related context, studies of patients with alexithymia may also be relevant [128]. These patients are traditionally classified as having difficulty recognizing, imagining, and thinking about emotions. However, it appears likely that, in addition to having problems in emotion-related cognition, these patients may also experience emotions rather differently. Some evidence for this comes from standard diagnostic questionnaires [e.g., Online Alexthymia Questionnaire (http://oaq.blogspot.com/)], including items such as 'I make decisions based on principles rather than gut feelings'; 'I sometimes experience confusing sensations in my body'; and 'For me sex is more a functional activity than it is an emotional one.' Indeed, one imaging study showed that these patients showed activity differing from control subjects in higher-order (i.e., prefrontal) regions, and in the insula cortex, but not in the amygdala [129]. HOT, when modified in the ways described in Box 3, provides a way of accounting for such alterations in subjective experiences.

Concluding Remarks

Taken together, these considerations suggest that HOT can be considered as a middle position between GWT and early sensory first-order local theories. GWT posits that consciousness is constitutively associated with high-level global cognition [8–10]. First-order local [87,90,130] theories suggest that no higher-level cognitive processes are constitutively required. HOT suggests that specific high-level processes are required, but such processes do not necessarily lead to global broadcasting and advantages in performance in simple tasks [13].

This middle position offers certain advantages. First, it accounts for why some powerful forms of unconscious processing are possible (e.g., blindsight). Second, it eliminates the cumbersome task faced by local theories of having to explain every distinct kind of phenomenology with a different first-order mechanism. Third, while GWT and HOT both propose a unified mechanism for consciousness, we argue that HOT has advantages accounting for how phenomenal experiences are assembled. Finally, we argue that HOT more readily accounts for the complexity of the phenomenology underlying not only our perceptions, but also our memories and emotions, and our awareness of our self, in both mental health and psychopathological conditions. Some issues for future work are listed in Outstanding Questions.

Acknowledgments

The authors thank William Chang for bibliographic assistance.

References

1. Chalmers, D. (1996) *The Conscious Mind*, Oxford University Press
2. Nagel, T. (1974) What is it like to be a bat? *Philos. Rev.* 83, 4435–4450
3. Rosenthal, D.M. (2005) *Consciousness and Mind*, Oxford University Press
4. Dretske, F. (1995) *Naturalizing the Mind*, MIT Press
5. Block, N. (2019) Empirical science meets higher order views of consciousness: reply to Hakwan Lau and Richard Brown. In *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness* (Pautz, A. and Stoljar, D., eds), pp. 199–213, MIT Press
6. Lamme, V.A. (2010) How neuroscience will change our view on consciousness. *Cogn. Neurosci.* 1, 204–220
7. Tye, M. (2000) *Consciousness, Color, and Content*, MIT Press
8. Baars, B.J. and Franklin, S. (2007) An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA. *Neural Netw.* 20, 955–961

Outstanding Questions

Can normal conscious experiences occur in the absence of first-order activity? While evidence exists that stimulation of PFC can change conscious experience, it is not clear whether this happens in the absence of distal impact on first-order sensory activity. This might be studied in future experiments by monitoring activity in first-order regions while stimulating higher-order regions at an intensity that would not lead to widespread network changes.

Taking the above question one step further, we can ask whether conscious experience can be preserved when the first-order regions are completely absent (because of a brain lesion or temporary inactivation). While such alterations would be expected to result in changes in higher-order states by reducing input signals to the higher-order network, future advances in methods may make it possible to stimulate higher-order regions in ways that mimic spatial and temporal properties of normal activity in the absence of a contribution of first-order states. With such methods, it might someday be possible to resolve current debates.

Do patients with damage to lower-order circuits typically associated with emotions (e.g., the amygdala) or patients with alexithymia, experience emotions differently from control subjects? If so, are these accounted for by high-level emotional processing at the conceptual level? For example, are schema involved, and does their activation pattern complete the subjective experience in the presence of trigger stimuli, even if body feedback is reduced or eliminated by damage to subcortical circuits?

How do we exactly delineate between the lower-order and higher-order networks? Out of convenience, we often define them anatomically. However, with respect to the input and output within a processing hierarchy, this may be more complicated. Does the higher-order network only involve prefrontal areas and, if so, which ones? What about areas of the temporal and parietal lobes? Given that any area that contributes inputs to the higher-order network is, in effect, lower order, some prefrontal areas may be lower order in the sense meant in HOT.

9. Dehaene, S. (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*, Penguin Books
10. Naccache, L. and Dehaene, S. (2007) Reportability and illusions of phenomenality in the light of the global neuronal workspace model. *Behav. Brain Sci.* 30, 518–520
11. Naccache, L. (2018) Why and how access consciousness can account for phenomenal consciousness. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373, 20170357
12. Rosenthal, D. (2004) Varieties of higher-order theory. In *Higher-Order Theories of Consciousness: An Anthology* (Gennaro, R.J., ed.), pp. 17–44, John Benjamins
13. Lau, H. and Rosenthal, D. (2011) Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373
14. Rosenthal, D. and Weisberg, J. (2008) Higher-order theories of consciousness. *Scholarpedia* 3, 4407
15. Armstrong, D.M. (1968) *A Materialist Theory of the Mind*, Routledge
16. Lycan, W.G. (2004) The superiority of Hop to HOT. In *Higher-Order Theories of Consciousness: An Anthology* (Gennaro, R.J., ed.), pp. 93–114, John Benjamins
17. Sauret, W. and Lycan, W.G. (2014) Attention and internal monitoring: a farewell to HOP. *Analysis* 74, 363–370
18. Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge University Press
19. Metzinger, T. (2003) *Being No One*, MIT Press
20. Cleeremans, A. (2011) The Radical Plasticity Thesis: how the brain learns to be conscious. *Front. Psychol.* 2, 86
21. Kriegel, U. (2009) *Subjective Consciousness: A Self-Representational Theory*, Oxford University Press
22. Brown, R. (2015) The HOROR theory of phenomenal consciousness. *Philos. Stud.* 172, 1783–1794
23. Lau, H. and Brown, R. (2019) The emperor's new phenomenology? The empirical case for conscious experience without first-order representations. In *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness* (Pautz, A. and Stoljar, D., eds), pp. 171–197, MIT Press
24. LeDoux, J.E. and Brown, R. (2017) A higher-order theory of emotional consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 114, E2016–E2025
25. Adams, F. and Shreve, C. (2016) What can synesthesia teach us about higher order theories of consciousness? *Symposion* 3, 251–257
26. Brinck, I. and Kirkeby-Hinrup, A. (2017) Change blindness in higher-order thought: misrepresentation or good enough? *J. Conscious. Stud.* 24, 50–73
27. Block, N. (2011) The higher-order approach to consciousness is defunct. *Analysis* 71, 419–431
28. Malach, R. (2011) Conscious perception and the frontal lobes: comment on Lau and Rosenthal. *Trends Cogn. Sci.* 15, 507 author reply 508–509
29. Coleman, S. (2018) The merits of higher-order thought theories. *Trans/Form/Ação* 41, 31–48
30. Rosenthal, D. (2018) Misrepresentation and mental appearance. *Trans/Form/Ação* 41, 49–74
31. Lau, H.C. and Passingham, R.E. (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18763–18768
32. Persaud, N. *et al.* (2011) Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *Neuroimage* 58, 605–611
33. Rounis, E. *et al.* (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* 1, 165–175
34. Ruby, E. *et al.* (2018) On a 'failed' attempt to manipulate visual metacognition with transcranial magnetic stimulation to prefrontal cortex. *Conscious. Cogn.* 62, 34–41
35. Fleming, S.M. *et al.* (2014) Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain* 137, 2811–2822
36. Lau, H. and Rosenthal, D. (2011) The higher-order view does not require consciously self-directed introspection: response to Malach. *Trends Cogn. Sci.* 15, 508–509
37. Odegaard, B. *et al.* (2017) Should a few null findings falsify prefrontal theories of conscious perception? *J. Neurosci.* 37, 9593–9602
38. Seager, W.E. (2004) A cold look at HOT theory. In *Higher-Order Theories of Consciousness: An Anthology* (Gennaro, R.J., ed.), pp. 255–275, John Benjamins
39. Jack, A.I. and Shallice, T. (2001) Introspective physicalism as an approach to the science of consciousness. *Cognition* 79, 161–196
40. Nelson, T.O. and Narens, L. (1994) Why investigate metacognition? In *Metacognition: Knowing about Knowing* (Metcalfe, J. and Shimamura, A.P., eds), pp. 1–25, The MIT Press
41. Dennett, D. (2015) Not just a fine trip down memory lane: comments on the essays on content and consciousness. In *Content and Consciousness Revisited: With Replies by Daniel Dennett* (Muñoz-Suárez, C. and De Brigard, F., eds), pp. 199–220, Springer International Publishing
42. Lau, H. (2019) Consciousness, metacognition, & perceptual reality monitoring. *PsyArXiv* 10 June
43. Damasio, A. (2010) *Self Comes to Mind: Constructing the Conscious Brain*, Pantheon Books
44. Baker, L.R. (2013) *Naturalism and the First-Person Perspective*, Oxford University Press
45. Panksepp, J. (2012) *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotion*, W.W. Norton & Company
46. Lewis, M. (2013) *The Rise of Consciousness and the Development of Emotional Life*, The Guilford Press
47. Rosenthal, D. (2002) Explaining consciousness. In *Philosophy of Mind: Classical and Contemporary Readings* (Chalmers, D.J., ed.), pp. 406–417, Oxford University Press
48. LeDoux, J.E. (2015) *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*, Viking
49. LeDoux, J. (2019) *The Deep History of Ourselves: How Ancient Microbes Became Conscious Brains*, Viking
50. Shea, N. *et al.* (2014) Supra-personal cognitive control and metacognition. *Trends Cogn. Sci.* 18, 186–193
51. Tomasello, M. and Rakoczy, H. (2003) What makes human cognition unique? From individual to shared to collective intentionality. *Mind Lang.* 18, 121–147
52. MacLean, E.L. (2016) Unraveling the evolution of uniquely human cognition. *Proc. Natl. Acad. Sci. U. S. A.* 113, 6348–6354
53. Tulving, E. (2005) Episodic memory and autonoesis: uniquely human? In *The Missing Link in Cognition* (Terrace, H.S. and Metcalfe, J., eds), pp. 4–56, Oxford University Press
54. LeDoux, J.E. (2017) Semantics, surplus meaning, and the science of fear. *Trends Cogn. Sci.* 21, 303–306
55. Hampton, R.R. (2001) Rhesus monkeys know when they remember. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5359–5362
56. Templer, V.L. *et al.* (2018) Rhesus monkeys metacognitively monitor memories of the order of events. *Sci. Rep.* 8, 11541
57. Panagiotaropoulos, T.I. *et al.* (2012) Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* 74, 924–935
58. Vandekerckhove, M. and Panksepp, J. (2009) The flow of anoetic to noetic and autonoetic consciousness: a vision of unknowing (anoetic) and knowing (noetic) consciousness in the remembrance of things past and imagined futures. *Conscious. Cogn.* 18, 1018–1028
59. Penn, D.C. *et al.* (2008) Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130, discussion 130–178
60. Shettleworth, S.J. (2010) Clever animals and killjoy explanations in comparative psychology. *Trends Cogn. Sci.* 14, 477–481
61. Heyes, C. (2015) Animal mindreading: what's the problem? *Psychon. Bull. Rev.* 22, 313–327
62. Gray, J.A. (2004) *Consciousness: Creeping Up on the Hard Problem*, Oxford University Press
63. Griffin, D. (1977) *The Question of Animal Awareness*, Rockefeller University Press
64. Bekoff, M. (2007) *The Emotional Lives of Animals: A Leading Scientist Explores Animal Joy, Sorrow, and Empathy – and Why They Matter*, New World Library
65. Burghardt, G.M. (2004) Ground rules for dealing with anthropomorphism. *Nature* 430, 15
66. de Waal, F. (2016) *What I learned from tickling apes*, New York Times 8 April

67. de Waal, F.B.M. (1999) Anthropomorphism and anthropodenial: consistency in our thinking about humans and other animals. *Philos. Top.* 27, 255–280
68. Maniscalco, B. and Lau, H. (2012) A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* 21, 422–430
69. Fleming, S.M. and Lau, H.C. (2014) How to measure metacognition. *Front. Hum. Neurosci.* 8, 443
70. Lau, H.C. and Passingham, R.E. (2007) Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* 27, 5805–5811
71. Shekhar, M. and Rahnev, D. (2018) Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *J. Neurosci.* 38, 5078–5087
72. Rosenthal, D. (2019) Consciousness and confidence. *Neuropsychologia* 128, 255–265
73. Dehaene, S. *et al.* (2017) What is consciousness, and could machines have it? *Science* 358, 486–492
74. Pearson, J. *et al.* (2015) Mental imagery: functional mechanisms and clinical applications. *Trends Cogn. Sci.* 19, 590–602
75. Berger, G.H. and Gaunitz, S.C. (1979) Self-rated imagery and encoding strategies in visual memory. *Br. J. Psychol.* 70, 21–24
76. Dehaene, S. *et al.* (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211
77. Sperling, G. (1960) The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29
78. Block, N. (2008) Consciousness and cognitive access. *Proc. Aristot. Soc.* 108, 289–317
79. Tsuchiya, N. *et al.* (2015) No-report paradigms: extracting the true neural correlates of consciousness. *Trends Cogn. Sci.* 19, 757–770
80. Block, N. (2014) Rich conscious perception outside focal attention. *Trends Cogn. Sci.* 18, 445–447
81. Kouider, S. *et al.* (2007) Cerebral bases of subliminal and supraliminal priming during reading. *Cereb. Cortex* 17, 2019–2029
82. Lumer, E.D. and Rees, G. (1999) Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1669–1673
83. Tse, P.U. *et al.* (2005) Visibility, visual awareness, and visual masking of simple unattended targets are confined to areas in the occipital cortex beyond human V1/V2. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17178–17183
84. Vidal, J.R. *et al.* (2014) Intracranial spectral amplitude dynamics of perceptual suppression in fronto-insular, occipito-temporal, and primary visual cortex. *Front. Psychol.* 5, 1545
85. Kapoor, V. *et al.* (2018) Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Commun. Biol.* 1, 215
86. Prinz, J.J. (2012) *The Conscious Brain: How Attention Engenders Experience*, Oxford University Press
87. Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–499 discussion 499–548
88. Macknik, S.L. and Livingstone, M.S. (1998) Neuronal correlates of visibility and invisibility in the primate visual system. *Nat. Neurosci.* 1, 144–149
89. Moutard, C. *et al.* (2015) Spontaneous fluctuations and nonlinear ignitions: two dynamic faces of cortical recurrent loops. *Neuron* 88, 194–206
90. Zeki, S. (2001) Localization and globalization in conscious vision. *Annu. Rev. Neurosci.* 24, 57–86
91. Silvanto, J. (2015) Why is ‘blindsight’ blind? A new perspective on primary visual cortex, recurrent activity and visual awareness. *Conscious. Cogn.* 32, 15–32
92. LeDoux, J. and Daw, N.D. (2018) Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. *Nat. Rev. Neurosci.* 19, 269–282
93. Weiskrantz, L. (1997) *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford University Press
94. Schmid, M.C. *et al.* (2010) Blindsight depends on the lateral geniculate nucleus. *Nature* 466, 373–377
95. Brown, R. (2014) Consciousness doesn’t overflow cognition. *Front. Psychol.* 5, 1399
96. Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
97. Nee, D.E. and D’Esposito, M. (2018) The representational basis of working memory. *Curr. Top. Behav. Neurosci.* 37, 213–230
98. D’Esposito, M. and Postle, B.R. (2015) The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142
99. Knotts, J.D. *et al.* (2018) Neuroscience: the key to consciousness may not be under the streetlight. *Curr. Biol.* 28, R749–R752
100. Sebastián, M.Á. (2013) Not a HOT dream. In *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience* (Brown, R., ed.), pp. 415–432, Springer
101. Carhart–Harris, R.L. *et al.* (2014) The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Front. Hum. Neurosci.* 8, 20
102. Carlen, M. (2017) What constitutes the prefrontal cortex? *Science* 358, 478–482
103. Teffer, K. and Semendeferi, K. (2012) Human prefrontal cortex: evolution, development, and pathology. *Prog. Brain Res.* 195, 191–218
104. Barbas, H. and Pandya, D.N. (1989) Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 286, 353–375
105. Koechlin, E. *et al.* (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302, 1181–1185
106. Passingham, R.E. and Wise, S.P. (2012) *The Neurobiology of the Prefrontal Cortex*, Oxford University Press
107. Yeterian, E.H. *et al.* (2012) The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81
108. Preuss, T.M. (2011) The human brain: rewired and running hot. *Ann. N. Y. Acad. Sci.* 1225, E182–E191
109. Koch, C. (2018) What is consciousness? *Nature* 557, S8–S12
110. Barrett, L.F. (2017) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* 12, 1–23
111. Clore, G.L. and Ortony, A. (2013) Psychological construction in the OCC model of emotion. *Emot. Rev.* 5, 335–343
112. LeDoux, J.E. (1996) *The Emotional Brain*, Simon and Schuster
113. LeDoux, J.E. (2008) Emotional colouration of consciousness: how feelings come about. In *Frontiers of Consciousness: Chichele Lectures* (Weiskrantz, L. and Davies, M., eds), pp. 69–130, Oxford University Press
114. Beck, A.T. (1976) *Cognitive Therapy and the Emotional Disorders*, International Universites Press
115. Hofmann, S.G. and Doan, S.N. (2018) *The Social Foundations of Emotion: Developmental, Cultural, and Clinical Dimensions*, American Psychological Association
116. Leahy, R.L. (2015) *Emotional Schema Therapy*, Guilford Press
117. LeDoux, J. *et al.* (2018) *Know Thyself: Well-Being and Subjective Experience*, The Dana Foundation
118. LeDoux, J.E. and Pine, D.S. (2016) Using neuroscience to help understand fear and anxiety: a two-system framework. *Am. J. Psychiatry* 173, 1083–1093
119. Miller, G. (2010) Is pharma running out of brainy ideas? *Science* 329, 502–504
120. Fanselow, M.S. and Pennington, Z.T. (2018) A return to the psychiatric dark ages with a two-system framework for fear. *Behav. Res. Ther.* 100, 24–29
121. Pine, D.S. and LeDoux, J.E. (2017) Elevating the role of subjective experience in the clinic: response to Fanselow and Pennington. *Am. J. Psychiatry* 174, 1121–1122
122. Tamietto, M. and de Gelder, B. (2010) Neural bases of the nonconscious perception of emotional signals. *Nat. Rev. Neurosci.* 11, 697–709
123. Ohman, A. (2002) Automaticity and the amygdala: nonconscious responses to emotional faces. *Curr. Dir. Psychol. Sci.* 11, 62–66
124. Bornemann, B. *et al.* (2012) Can you feel what you do not see? Using internal feedback to detect briefly presented emotional stimuli. *Int. J. Psychophysiol.* 85, 116–124

125. Inman, C.S. *et al.* (2018) Human amygdala stimulation effects on emotion physiology and emotional experience. *Neuropsychologia* Published online March 15, 2018. https://doi.org/10.1016/j.neuropsychologia.2018.03.019
126. Anderson, A.K. and Phelps, E.A. (2002) Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J. Cogn. Neurosci.* 14, 709–720
127. Feinstein, J.S. *et al.* (2013) Fear and panic in humans with bilateral amygdala damage. *Nat. Neurosci.* 16, 270–272
128. Silani, G. *et al.* (2008) Levels of emotional awareness and autism: an fMRI study. *Soc. Neurosci.* 3, 97–112
129. Bernhardt, B.C. *et al.* (2014) Selective disruption of sociocognitive structural brain networks in autism and alexithymia. *Cereb. Cortex* 24, 3258–3267
130. Fisch, L. *et al.* (2009) Neural 'ignition': enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron* 64, 562–574
131. Shea, N. and Frith, C.D. (2019) The global workspace needs metacognition. *Trends Cogn. Sci.* 23, 560–571
132. Macmillan, N.A. and Creelman, C.D. (2004) *Detection Theory: A User's Guide* (2nd edn), Lawrence Erlbaum Associates
133. Peters, M.A. and Lau, H. (2015) Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *Elife* 4, e09651
134. Persuh, M. *et al.* (2018) Working memory and consciousness: the current state of play. *Front. Hum. Neurosci.* 12, 78
135. Seth, A.K. (2019) From unconscious inference to the beholder's share: predictive perception and human experience. *Eur. Rev.* Published online June 14, 2019. https://doi.org/10.1017/S1062798719000061
136. Odegaard, B. *et al.* (2018) Inflation versus filling-in: why we feel we see more than we actually do in peripheral vision. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 373, 20170345
137. Komatsu, H. (2006) The neural mechanisms of perceptual filling-in. *Nat. Rev. Neurosci.* 7, 220–231
138. Li, M.K. *et al.* (2018) An investigation of detection biases in the unattended periphery during simulated driving. *Atten. Percept. Psychophysiol.* 80, 1325–1332
139. Postle, B.R. (2006) Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38