Cell
PRESS

Empirical support for higher-order theories of conscious awareness

Hakwan Lau[1,2] and David Rosenthal[3]

[1] Department of Psychology, Columbia University, 355D Schermerhorn Ext, 1190 Amsterdam Avenue MC: 5501, New York, NY 10027, USA
[2] Donders Institute for Brain, Cognition and Behavior, Radboud University, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
[3] Philosophy Program and Concentration in Cognitive Science, Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016-4309, USA

Higher-order theories of consciousness argue that conscious awareness crucially depends on higher-order mental representations that represent oneself as being in particular mental states. These theories have featured prominently in recent debates on conscious awareness. We provide new leverage on these debates by reviewing the empirical evidence in support of the higher-order view. We focus on evidence that distinguishes the higher-order view from its alternatives, such as the first-order, global workspace and recurrent visual processing theories. We defend the higher-order view against several major criticisms, such as prefrontal activity reflects attention but not awareness, and prefrontal lesion does not abolish awareness. Although the higher-order approach originated in philosophical discussions, we show that it is testable and has received substantial empirical support.

The empirical implications of higher-order theories of conscious awareness

Although by no means unanimously supported, the higher-order view of conscious awareness (see Glossary) is popular among philosophers and scientists alike [1–10]. It is sharply contrasted with first-order theories [11–13], which hold that a mental state being conscious is determined solely by the neural or representational character of the perceptual state (i.e. first-order states). For instance on a first-order view the conscious experience of seeing a face is determined by the representation of the face alone. Higher-order theories, by contrast, suggest that first-order representations alone are not sufficient, because some first-order representations occur outside conscious awareness. Conscious awareness crucially depends on higher-order representations, specifically mental states that represent oneself as being in the relevant first-order mental states. The basic motivation is that no mental state is conscious if one is wholly unaware of being in such a mental state.

The roots of the higher-order view can be traced back to Locke and Kant [14–16], and there are currently several variants of the higher-order view (Table 1). Despite the differences among these variants [16], they hold in common that a mere change in the higher-order representation

or process is sufficient to lead to a change in subjective awareness, even if all first-order representations remain the same. This distinguishes them sharply from first-order theories.

Despite many challenges from within philosophy (e.g. recently [17]), the view has been ably defended ([1,5,6,8]; recently [18,19]). In this article we focus on exciting new evidence emerging from cognitive neuroscience in support of the higher-order view that gives the view substantial empirical credibility.

The link between the higher-order view and neuroscientific data stems from taking the view that first-order perceptual representations depend on neural activity in early sensory regions, whereas higher-order representations depend on neural activity mainly in prefrontal (and

---

Glossary

Conscious awareness (sometimes **consciousness** or **awareness** for short): a term that has provoked much controversy. We use it here to refer to perceptual processes that occur with subjective experience, of which we are aware and about which we can report under normal circumstances. This contrasts with perceptual processes that occur without subjectivity, of which we are unaware and about which we cannot report. The term 'conscious awareness' can also apply to thoughts and volitional states of which one is subjectively aware, although these are not the main focus here.

Metacognition: cognition that is about another cognitive process as opposed to about objects in the world. In this article we use it mainly to refer to sensory metacognition: a cognitive process that concerns the quality or efficacy of a perceptual process. The capacity of sensory metacognition is sometimes empirically assessed by the correspondence between subjective report and task performance: how closely are they associated with each other on a trial-by-trial basis.

Qualitative character of conscious perception (or sometimes **qualitative character** or **mental quality**): the subjective experiential character associated with a perceptual event or process: what is it like for the subject [93].

Subjective ratings/report (of awareness): reports that are either verbal or expressed via 'commentary keys' [3] by the subject, often immediately after performing a perceptual task. The purpose of these ratings is to indicate whether a relevant stimulus property has been consciously perceived as opposed to subjects merely 'guessing' in the perceptual task. A similar measure relies on confidence ratings in a discrimination task. Although the frequency of high subjective ratings is often correlated with task performance, the two are distinct and sometimes dissociate [92]. Because a verbal report expresses a thought about the reported state, higher-order-thought theory provides an explanation of the widely accepted correlation between conscious awareness and reportability [1].

Task performance: in this article this term mostly refers to performance level in a perceptual task, such as discrimination between two stimulus alternatives (e.g. a face vs a house, a gabor patch vs noise or a sound on the left vs a sound on the right). Task performance is often measured in terms of accuracy (% correct) or the signal detection theoretic measure d', which is an estimate of the underlying processing capacity in terms of signal-to-noise ratio.

Corresponding authors: Lau, H. (hakwan@gmail.com); Rosenthal, D. (davidrosenthal@nyu.edu).

Table 1. Varieties of higher-order approaches to conscious awareness

Name of variant	Essential higher-order mechanism	How does awareness arise?	Key references
Higher-order thought (hot) theory	A thought-like mental state. (Note: As with most higher-order theories, this mental state typically is not conscious)	When a higher-order thought-like representation results in one's being aware of a first-order representation	[1]; also endorsed by [3,101]
Higher-order perception theory	A perception-like mental state	When a higher-order perception-like mental state 'senses' a first-order representation	[4,5]
Same-order theory	A state's referring to itself	When a mental state with perceptual content represents itself	[6,10,102]
Dispositional higher-order thought theory	A disposition to have a higher-order thought-like state	When a first-order representation becomes disposed to be represented by a higher-order thought-like mental state	[8]
Higher-order statistical inference view	A cognitive process akin to perceptual decision making	When a first-order representation is 'judged' by a higher-order decision process to constitute a statistically reliable perceptual signal. (Note that a poorly activated higher-order process can still make faulty decisions, including false alarms. Thus low activity in the higher-order process does not necessarily mean low level of awareness.)	[2,55]
Radical plasticity thesis	A learning-based representational redescription process	When the higher-order process 'learns' that a reliable first-order representation exists within the system. Because learning results in a thought-like representation, this is akin to the higher-order thought theory.	[9,103]

parietal) cortex in humans and other primates. By itself, the higher-order view does not specify this neuroanatomical detail, but this interpretation is accepted by many authors [2,6] including first-order theorists [12,13]. We adopt this interpretation throughout this article.

Distinguishing the higher-order view from alternatives
In this article we consider three major alternatives to the higher-order view: neuronal global workspace theory [20–22], information integration theory [23] and the first-order view [11–13,24–26]. These four major theories make different empirical predictions (Table 2).

According to the neuronal global workspace theory [20–22], neurons with long-range connections in the prefrontal and parietal areas form a network (i.e. workspace) for conscious processing. As with the higher-order view, the global workspace theory predicts that awareness is determined by prefrontal and parietal activity. However, there is a crucial difference: according to the neuronal global workspace theory, the awareness-related activity in the

prefrontal and parietal areas is associated with essential behavioral functions, such as flexible control of behavior, cognitive control and ability to perform various tasks [20–22]. So, on this view, the potential for good performance, in both perceptual and higher cognitive tasks, is crucial for a representation to be conscious.

The higher-order view, by contrast, is neutral about whether conscious awareness adds significant utility or immediate impact on behavior and task performance [1,27]. This is because the view assumes that task performance in most perceptual and cognitive tasks depends mainly on first-order rather than higher-order representations. Because conscious awareness can differ even if all first-order representations remain completely unchanged, such awareness itself might serve little function [1,27].

Information integration theory [28] holds that awareness depends on complex patterns of connections between neurons that allow for the representation of many different possible informational states by a single 'core' system. So far, this core system has not been strictly identified with a

Table 2. Main empirical predictions by major theories of conscious awareness

Theory	Predicts changes in relevant prefrontal (and parietal) activity is sufficient for changes in awareness	Associates awareness with performance capacity in perceptual tasks	Associates awareness with higher cognitive functions (e.g. cognitive control)
Higher-order view	Yes (although it does not predict that changes in 'any' prefrontal or parietal activity is sufficient; higher-order representations probably constitute only a small subset of such activity)	No	No
Neuronal global workspace theory	Yes	Yes (although one alternative is to adopt a dual channel variant of the global workspace view; see Figure 1b)	Yes
Information integration theory	Yes (assuming some 'relevant' activity in prefrontal and parietal cortices reflects the 'core' system)	No	Maybe (if the task requires very complex and context-dependent stimulus-response mapping)
First-order theories (including Recurrent processing view)	No	Yes (although not always a consequence of the theory, first-order theorists have generally assumed this to be the case [11–13,25,26])	No

specific anatomical structure. However, during wakefulness, the core is likely to depend, at least in part, on activity in the prefrontal and parietal areas where connections are complex. Regarding functions of conscious awareness, information integration theory does not hold that unconscious information processing (i.e. outside the core) is necessarily associated with low task performance. Unconscious processing can be strong and reliable. What it lacks is complexity, that is, the ability to represent many different possibilities within a single system. Thus certain higher-level cognitive tasks that require high context sensitivity, such as complex stimulus-response mapping, might require awareness (G. Tononi, personal communication).

The first-order view maintains that conscious awareness is determined by early sensory activity alone, independently of higher-order representations. Thus the crucial difference between the first-order and higher-order views is that the latter but not the former predict that conscious awareness is determined at least in part by prefrontal and parietal activity.

Therefore, in the group of first-order theorists we include here not only those commonly labeled as such in the philosophical literature [11] but also theorists, such as Ned Block [12,13], who hold that the phenomenal qualities of awareness depend on the biological substrate rather than merely the content of the first-order representations, as well as scientists who hold that visual awareness depends on activity in content-specific regions of the extrastriate cortex [24,25] or on feedback from these regions to the primary visual cortex [26].

Most first-order theories explain the difference between awareness and unawareness by positing that the latter is associated with weak information [29] or with representations in alternative (e.g. subcortical) sensory pathways. Thus this view suggests that conscious and unconscious processing will have different functional consequences. Whereas some first-order theorists hold that some higher

cognitive functions can be performed without conscious awareness [30–32], most first-order theorists take a difference in perceptual task performance (e.g. 'hits' vs 'misses' in a detection task) as evidence for a difference in conscious awareness [12,13,25,26,33]. In other words, they typically associate task performance with awareness.

Conscious awareness is associated with specific prefrontal mechanisms

Numerous neuroimaging studies have identified activations in prefrontal and parietal cortices in association with conscious awareness [22,34]. On the face of it, these findings support the higher-order, global workspace and information integration views over the first-order view.

The higher-order view can be distinguished from global workspace theory by conditions in which task performance is matched but conscious awareness differs. One study [35] used visual masking to create conditions in which the subjects' ability to discriminate between two figures remained the same; however, subjects claimed to be consciously aware more frequently of the identity of the figure (i.e. less frequently guessing) in one condition than in the other. This higher level of awareness was associated with activity in dorsolateral prefrontal cortex (Figure 1a). Transcranial magnetic stimulation (TMS) applied to this region also changed subjective reports of awareness without impairing task performance [36]. These results point toward the higher-order view rather than toward global workspace theory because global workspace theory suggests that activity in prefrontal cortex reflects not only conscious awareness but also the capacity for superior task performance.

One way to account for a change in awareness with constant task performance is to stipulate the existence of an independent, parallel, and unconscious channel of processing [21], which contributes to task performance but not awareness. Thus a change in balance between conscious
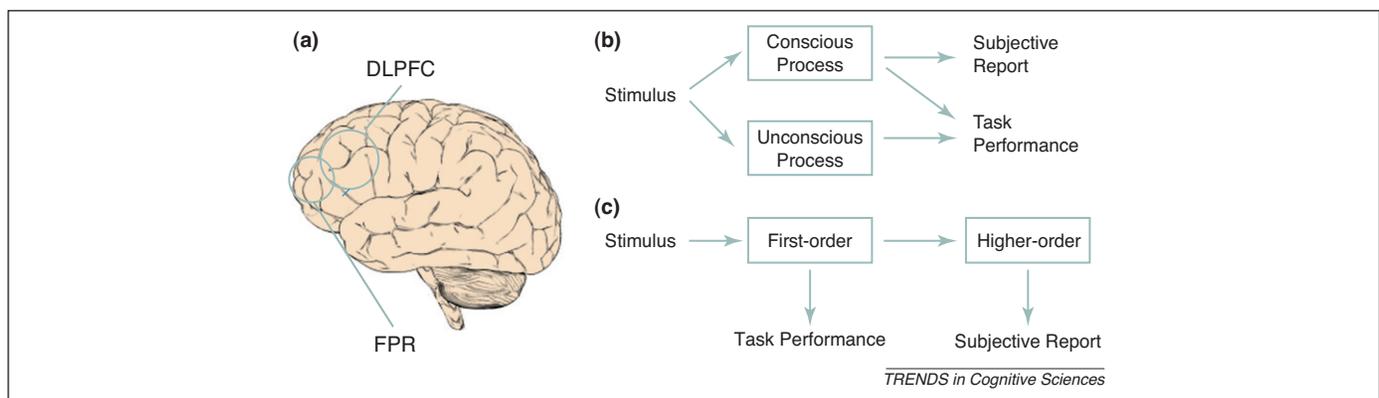
Figure 1. Prefrontal activity and late-stage metacognitive processes in a hierarchical system. **(a)** The dorsolateral prefrontal cortex (DLPFC; Brodmann's areas 46 and 9) has been implicated in numerous studies of visual awareness [22,34,56], including studies in which task performance was controlled for [35,36,68]. It has also been suggested that this region is involved in perceptual decision-making [94] and sensory metacognition [36]. Individual structural differences in this area, as well as in the neighboring frontal polar region (FPR; Brodmann's area 10), are known to reflect variations in the capacity of sensory metacognition [95] (i.e. how well one can place confidence ratings to distinguish between one's own correct and incorrect visual judgments). **(b)** A dual-process model can account for a change in visual awareness in the absence of a change in task performance [35,36,68]. To keep overall task performance constant (which depend on both channels), a change in the level of activity in the 'conscious' channel needs to be accompanied by a change in activity in the opposite direction in the 'unconscious' channel. With this model, one plausible interpretation is that activity in the conscious channel is reflected by activity in the prefrontal and perhaps other cortical areas, whereas the unconscious channel is reflected by activity in subcortical areas [21]. **(c)** More in accord with the higher-order view is a hierarchical model where the first-stage process that drives task performance is not itself conscious, and a second-stage process 'monitors' or responds to changes in the first stage to determine what enters conscious awareness. This is compatible with recent work in connectionist [9] and diffusion [37] modeling on sensory metacognition. In a study [38] that directly compared the hierarchical model against the dual channel model, it was found that the former produced better fits to data. Taken together, we suggest that the prefrontal activity implicated in studies of awareness reflects late-stage metacognitive processing in the hierarchical model.

Box 1. Conscious perception as a decision process

There is a long and influential tradition in psychophysics that treats perception as a decision-making process [76]. Given the ubiquitous presence of noise in all physiological systems, the brain must 'decide' whether a certain amount of neural activity represents a reliable signal or merely baseline fluctuation. This can be achieved by setting a criterion to distinguish signal from noise [76–78] or by other higher-order interpretative mechanisms that read out the content from the noisy neuronal code [79].

One might argue that such decision-making mechanisms are postperceptual and are only for generating responses within the context of psychological tasks. However, a high level of sensory activity cannot by itself constitute reliable perception, since it could be that the baseline noise level is also high, and the signal must be interpreted within that context. Therefore, these decision mechanisms are integral to the perceptual process itself, unless such a process does not concern signal quality as in perhaps most cases of 'subliminal' perception.

Compared to subliminal perception, normal conscious perception seems typically to be associated with a relatively high quality of representation (although that might not always be the case, e.g. [3,35,36,68] and Figure 2). Conscious information can facilitate serial processing (e.g. conscious thinking and planning), which is notoriously intolerant to noise as pointed out by von Neumann [80].

Because of this, Dehaene [21] suggested that conscious awareness might be based on decision mechanisms that set a high criterion to select only signals of good quality. This can be achieved by the metacognitive mechanisms described in Figure 1.

The necessity to 'decide' what enters conscious awareness based on a high criterion might be especially clear in temporal contexts. For example, Libet famously argued that our awareness of an intention to produce a spontaneous movement significantly lags behind the onset of the corresponding motor preparation signal in the brain [81]. Recent models [82,83] stipulate that this is because we only become aware of the motor preparation signal after it ramps up to reach a certain level of intensity. Nikolov *et al.* [83] specifically gave an explanation of why this high criterion strategy can be statistically advantageous.

These authors [83] also applied the model to another phenomenon known as attentional prior entry [84,85], whereby attention to an object can make subjects consciously perceive it as arising earlier than unattended objects that are presented simultaneously. One interesting fact about attentional prior entry is that researchers have failed so far to find early sensory temporal signatures that can account for the shifts in perceptual onset due to attention [84,85]. Thus, the phenomenon is likely to depend on higher-order mechanisms [83,84].

and unconscious channels can account for a change in awareness while leaving task performance unchanged (Figure 1b). However, recent computational modeling work [9,37,38] suggests that unconscious and conscious processes probably form a serial hierarchy rather than two parallel channels (Figure 1c). In line with the higher-order view, these computational models suggest that conscious awareness depends on a late-stage monitoring (i.e. metacognitive) process, which determines whether a first-order state represents reliable information (Box 1). This is likely to be the mechanism by which awareness of the first-order states is generated.

The argument from the possible function of conscious awareness

Although studies of subliminal priming demonstrate that unconscious information can influence behavior, such influence has traditionally been thought to be minimal [39,40]. One difficulty in assessing this claim is that the usual method for rendering a stimulus subjectively invisible involves impairing the perceptual signal itself until task performance is at chance [41]. Thus, when subjects perform cognitive tasks poorly with subjectively invisible stimuli, it is often difficult to know whether that is due to lack of conscious awareness, since it may instead be due to a degraded perceptual signal. Therefore, current research tends to underestimate the power of the unconscious [27,41]. Despite this methodological problem, recent studies have shown that unconscious information can facilitate or at least influence higher cognitive functions such as task preparation [42], self-motivation [43], pursuit of goals [44], learning [45,46], mental arithmetic [47], response inhibition [30,32], error detection [31], information processing across sensory modalities [48], as well as combining different features and objects to form a global 'percept' [49]. These findings suggest that awareness might not have 'special' utility [27,50], at least regarding its immediate impact on behavior [41,51]. Although this claim is counterintuitive and somewhat controversial [52,53], it is com-

patible with the predictions made by the higher-order view [27] (Table 2).

Note that even if conscious awareness turns out to be associated with certain specific higher cognitive functions this does not in itself falsify the higher-order view. One advantage of the higher-order view over its alternatives is that it does not make strong *a priori* assumptions regarding the possible functions of awareness, some of which are already known not to be true [30,42]. By itself, the higher-order view is not committed to the prediction that awareness has no function whatsoever; it is neutral with respect to this issue.

Empirically based criticisms of the higher-order view

It could be argued that the prefrontal activity that occurs with conscious awareness [22,34] reflects something other than awareness itself, perhaps attention or access to the perceptual information. Two studies that have controlled for attention [24,54] by requiring subjects to perform a task unrelated to the visual target in question would seem to be compatible with this argument. These studies found that visible targets (compared to targets rendered invisible by masking), when unattended, were associated with activity in the early visual areas but not in the prefrontal cortex. This may be taken to suggest that prefrontal activity reflects attention or access instead of conscious awareness. (For a related argument that prefrontal activity reflects postperceptual access or decision making see Box 1.)

However, there is reason to think that this is not the correct conclusion. Attention may indeed have been a confounding factor in some previous neuroimaging studies of conscious awareness. But, attention probably cannot account for the finding that TMS to the prefrontal cortex changes subjective reports of awareness [36]. It is unlikely that the reports in that study differed because TMS manipulated attention, since task performance was unaltered. Second, some higher-order theories do not hold that conscious awareness is invariably associated with increased prefrontal activity, for example the higher-order

Box 2. The myth that prefrontal damage does not affect visual perception

It is often claimed that damage to the prefrontal cortex, unlike damage to the primary visual cortex, does not abolish visual awareness [65]. Although this is true in many cases, it is unclear whether comparing prefrontal and early sensory lesions is fair or informative, since damage to the primary visual cortex abolishes most visual input to the rest of the cortex and so can be confounded with potential distal effects. On the other hand, the prefrontal cortex is a large and flexible central system [86]. It has been shown that if the damage to the prefrontal cortex is unilateral (which is typically the case), the undamaged hemisphere can dynamically 'take over' and contribute to functions normally carried out by the damaged side [87].

Although rare, the case of a patient with large bilateral damages to the lateral prefrontal cortex has been tested by Knight and colleagues [86], and the patient was generally unresponsive to external stimuli, as if conscious awareness was completely abolished. The same group also found that even when the lesion was unilateral, visual performance was impaired under a task that required simultaneous monitoring of visual information presented quickly on both hemifields [88]. Monkeys with large unilateral lesions to the frontal and parietal cortices are also known to behave as though they are blind [89]. These results are compatible with the higher-order view as well as the global workspace and information integration view.

One could argue that these cases [86,88,89] were due to impairments of endogenous attention, rather than of awareness *per se*.

However, a recent study has reported specific impairments in patients with smaller lesions to the prefrontal cortex (in particular the frontal polar region) [90]. It was found that visual discrimination performance was uniformly impaired under varying degrees of attentional cuing. Interestingly, the effect on subjective reports was more salient than the effect on task performance.

Still more specific results can be obtained with TMS. Rounis *et al.* [36] reported that TMS targeted at the dorsolateral prefrontal cortex both lowered subjective reports of visual awareness and impaired metacognitive ability. Again, the effect was salient in the subjective reports, whereas task performance was unimpaired. Even though this study involved bilateral stimulation, unilateral TMS to the right dorsolateral prefrontal cortex is also known to impair performance in change blindness [91].

The fact that the effects of prefrontal disturbance can be relatively specific to subjective reports may explain the paucity of evidence from animal lesion studies, which mainly rely on task performance measures. However, because subjective reports are more direct measures of visual awareness compared to task performance [92] (Figure 2), this is in agreement with the claim proposed by the higher-order view that the prefrontal cortex is crucial for conscious awareness, and in particular the subjective aspects (i.e. not only task performance).

statistical inference view [55] (Box 2). Moreover, subjects in the two attention-controlled studies [24,54] were required to respond to a different task and to ignore the targets; so there was no behavioral report to ascertain if the targets were consciously seen under lack of attention [56]. One might appeal to the results of introspection in doing these tasks and claim that awareness seems to occur even though little attention is paid to the targets. However, the reliability of such introspection is questionable [57,58],

especially because it has been shown that individuals consistently overestimate what they see in unattended locations [59] (Figure 2).

A related study [60] showed that, under rapid presentation of visual images, activity in prefrontal cortex was reduced. However, with rapid representation, it is probable that subjects consciously saw each image in less detail than in the control condition, when presentation was slower. Indeed, image categorization performance decreased un-
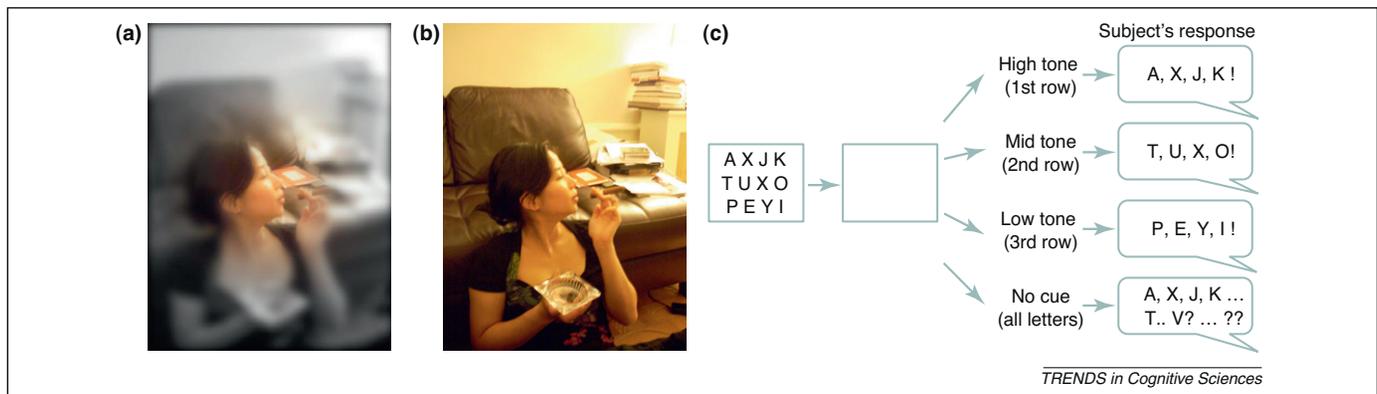
Figure 2. Inattentional inflation of subjective perception and Block's puzzle of 'phenomenological overflow'. **(a)** Peripheral vision is known to be associated with low spatial resolution [96] and low color sensitivity [97]. In everyday settings, the periphery also receives little attention compared to the foveal location. Change blindness and inattentional blindness studies have shown that we are poor at processing object information in unattended locations [98]. Based on these findings, one might expect our conscious visual experience to be similar to what is shown in **(a)**. However, there is a compelling subjective impression that peripheral vision is less impoverished: in particular, subjective vision is more similar to what is depicted in **(b)** rather than **(a)**. This results in a difficulty for the first-order view because early sensory processing probably does not support vivid details for the periphery [96,97]; how, then, can a first-order view explain the subjective impression of detail? On the higher-order view, one possibility is that under the lack of attention the higher-order system might mistake noisy early input as reliable signal: it might produce more false alarms and give higher confidence even when the task performance for unattended objects is not higher than for attended objects. This explanation receives support from Rahnev *et al.* [59] and Davidson *et al.* [99] who showed that although attention positively boosted task performance, it led to more conservative perceptual decisions (i.e. fewer hits and lower confidence) compared to inattention. This corresponds to the apparent richness of perception at unattended locations where qualitative character is subjectively inflated as compared to **(a)**. Furthermore, the subjective richness of qualitative character in peripheral vision could be due to memory from previous visual fixations at such locations. If so, the apparent richness of qualitative character is due to top-down 'filling in,' rather than detailed first-order representations at the moment of perception. This again fits well with the higher-order view. **(c)** Based on the results of Sperling postcue experiments [100], Block has argued that the richness of some details of our conscious experiences may overflow our ability to report or access cognitively [12]. In the standard Sperling paradigm, subjects were presented with an array of 12 letters. When cued for a particular row immediately after the letters disappeared, subjects were able to report the identities of the 4 letters in that row but not all 12 letters. Block argues that subjects consciously see the identity of all 12 letters so that higher-order reportability dissociates from awareness. However, this has been challenged by the recent empirical finding that subjects did not detect unusual symbols inserted in the uncued locations [54,58], which suggests that subjects only saw 12 letter-like objects in the array but not their identities and related details. On our view, subjects' sense that they consciously see the identities of all 12 letters is unwarranted, and is probably due to the phenomenon of 'inattentional inflation of subjective perception' described above. Because relatively little attention is paid to any particular letters when the display is present, it is probable that one mistakenly overestimates how much is consciously processed.

Box 3. The problem of division of phenomenal labor

Because the higher-order view suggests that both first-order and higher-order representations are involved in conscious perception, one might wonder what happens if the content of the two fails to match. For example, what would be the qualitative character of conscious perception if the first-order state represents green, but the higher-order state misrepresents one as being in a first-order state of seeing red? Some have argued the possibility of such cases is problematic for the higher-order view [12,74,75].

Indeed, higher-order representations may often diverge in relatively mild ways from first-order representations in respect of fineness of grain, for example with a first-order representation of a precise shade of red and a higher-order representation of a sensation of a more generic red. However, something similar can happen with late-stage misrepresentation on every view of conscious perception that stipulates a hierarchy of processing, including the global workspace view and Block's view of access consciousness [17]. In general, whenever there are two stages of processing, there is a theoretical possibility of mismatch. The question is which level of representation determines the qualitative character of conscious awareness when such mismatch occurs.

On Block's view [17], when there is such late-stage misrepresentation, the qualitative character of conscious perception is determined by the first-order representations in early sensory regions. By contrast, some proponents of higher-order theories (e.g. [1,18,19]) hold that in such cases the qualitative character of conscious perception is determined, as in all cases, solely by the way the higher-order representation makes one aware of one's first-order states. Other higher-order theorists hold that the qualitative character could be determined jointly by both first-order and higher-order representations [6,55]. It is unclear why such cases of mismatch result in greater difficulty for one theory over another, since the difference among these views is solely about which stage of processing determines qualitative character, early, late, or both. This is an empirical question that can be addressed by future experiments that directly test whether the qualitative character of conscious perception can be predicted entirely from prefrontal activity, early sensory activity alone, or both.

der rapid presentation [60]. Moreover, the prefrontal area identified in that study is likely to be related to self-related processing, and is, in any case, distinct from the dorsolateral prefrontal area that is hypothesized to be related to the type of higher-order awareness relevant here [35]. Higher-order representations that support conscious awareness might depend on specific mechanisms (Table 1; Box 1) and must be distinguished from other kinds of high-level cognitive processes.

The last point is related to the argument that children and animals presumably have conscious awareness, even though their prefrontal cortex might be underdeveloped [12,13]. However, it is an open question how fine-grained conscious awareness is in children and animals, and less fine-grained awareness might not require a fully developed prefrontal cortex. Also, although children and animals might lack self-referential conscious thinking [13], conscious awareness might not require that. On the higher-order view, self-referential conscious thinking results only when the higher-order representations are themselves conscious [1]; this requires third-order representations, which are likely a subsequent development, and so may be absent in small children and animals. Higher-order representations might be better reflected by sensory metacognition, for instance the ability to rate confidence appropriately to distinguish between correct and incorrect perception rather than by conscious self-referential thinking. Signs of sensory metacognition have been observed in animals such as rats and dolphins [61–63].

Related to this is a recent study that has shown that in some cases when subjects reported not seeing a visual target they were nonetheless able to demonstrate sensory metacognition [64]. This could be taken to argue against the higher-order view because it shows that sensory metacognition and awareness dissociate [22]. However, even though sensory metacognition and awareness are related, a specific kind of higher-order representation is required on the higher-order view for conscious awareness to occur (Table 1). It is entirely compatible with the higher-order view that subjects do not consciously perceive when they have a higher-order representation of *not* perceiving, and this accounts for the metacognitive ability [64].

Finally, it has been argued against the higher-order view that damage to prefrontal cortex does not affect visual awareness [65]. However, there is now ample empirical evidence (Box 2) that convincingly challenges this argument.

Suggestive evidence based on clinical disorders

Clinical cases are often useful in theoretical debates. Consider a neurological condition that might at first glance seem to tell against the higher-order view. In blindsight [3,66], visual awareness is abolished even though above-chance task performance is retained. Because blindsight is typically due to damage to the primary visual cortex, one might conclude that blindsight supports the first-order rather than the higher-order view. However, damage to an area often has distal effects, and it has been shown that the lack of conscious awareness in blindsight could be associated with reduced activity in the prefrontal cortex [67], even when task performance was controlled for [68]. Compatible with these findings, a recent computational model accounts for blindsight in terms of the effect of the early sensory lesion on higher-order representations [9].

Many other clinical cases of disturbance in conscious awareness are also evidently compatible with the higher-order view, since they have prefrontal origins. Schizophrenia is associated with disorders of the dopamine system that projects widely to the prefrontal cortex [69]. It is possible that the disturbed functions of the prefrontal cortex could at least partially explain the hallucinations experienced by these patients. Parkinson's patients who are treated with dopaminergic drugs (e.g. Levodopa) often develop visual hallucinations [70]. In presurgical epileptics, direct cortical stimulation to prefrontal cortex can also lead to visual hallucinations [71]. Thus conscious visual awareness, albeit hallucinatory, evidently can result from activity in prefrontal cortex. However, in these cases we cannot rule out that feedback to early sensory areas could be the ultimate cause of awareness (or alternatively, it could be a byproduct that is not itself crucial for conscious awareness).

Also relevant, and perhaps more decisive, are rare cases of Charles Bonnet syndrome that, unlike conventional cases of the disorder, involve damage to the primary visual

area [72,73]. These patients are cognitively cogent but experience vivid visual hallucinations. Although the damage does not include the entire visual cortex, these cases challenge those first-order views, which hold that feedback to primary visual areas is crucial for conscious visual awareness [26]. If conscious experience can exist in the absence of first-order representations, the qualitative character of conscious awareness might depend entirely on higher-order representations. Thus, these cases are also highly relevant to the problem raised by some [12,74,75] about the effect on conscious awareness when the first-order and higher-order representations do not match each other in content (Box 3).

Concluding remarks

We have reviewed empirical evidence that supports the higher-order view of conscious awareness and addressed empirically based challenges to the view. We focused on cases in which awareness differed despite matched task performance because these cases most crucially distinguish the higher-order view from its alternatives. Because such matching is experimentally difficult to achieve, the most decisive evidence is only now beginning to emerge. Our arguments are therefore not conclusive, and future experiments might challenge our conclusions (Box 4). Nevertheless, we hope that the preceding discussion convincingly demonstrates that, under a well-founded view about neural implementation, the higher-order view is supported by the currently available empirical evidence.

References

1 Rosenthal, D.M. (2005) *Consciousness and Mind*, Clarendon Press
2 Lau, H. (2010) Theoretical motivations for investigating the neural correlates of consciousness. *WIREs Cognit. Sci.* 2, 1–7
3 Weiskrantz, L. (1997) *Consciousness Lost and Found: A Neuropsychological Exploration*, Oxford University Press
4 Armstrong, D.M. (1980) What is consciousness? In *The Nature of Mind* (Armstrong, D.M., ed.), pp. 55–67, University of Queensland Press
5 Lycan, W.G. (1996) *Consciousness and Experience*, MIT Press
6 Kriegel, U. (2009) *Subjective Consciousness: A Self-Representational Theory*, Oxford University Press
7 Dienes, Z. (2008) Subjective measures of unconscious knowledge. *Prog. Brain Res.* 168, 49–64
8 Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge University Press
9 Pasquali, A. *et al.* (2010) Know thyself: metacognitive networks and measures of consciousness. *Cognition* 117, 182–190
10 Gennaro, R.J. (2011) *The Consciousness Paradox: Consciousness, Concepts and Higher-Order Thoughts*, MIT Press
11 Dretske, F.I. (1995) *Naturalizing the Mind*, MIT Press
12 Block, N. (2007) Consciousness, accessibility and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–548
13 Block, N. (2009) Comparing the major theories of consciousness. In *The Cognitive Neurosciences IV* (Gazzaniga, M., ed.), MIT Press
14 Locke, J. (1975/1700). In *An Essay concerning Human Understanding* (Nidditch, P.H., ed.), Oxford University Press
15 Kant, I. (1998/1787) *Critique of Pure Reason* (Guyer, P. and Wood, A.W., trs and eds), Cambridge University Press
16 Rosenthal, D.M. (2004) Varieties of higher-order theory. In *Higher-Order Theories of Consciousness* (Gennaro, R.J., ed.), pp. 19–44, John Benjamins Publishers
17 Block, N. (2011) The higher order approach to consciousness is defunct. *Analysis* 71, DOI: 10.1093/analys/anr037
18 Rosenthal, D.M. (2011) Exaggerated reports: reply to Block. *Analysis* 71, DOI: 10.1093/analys/anr039
19 Weisberg, J. (2011) Abusing the notion of what-it's-like-ness: a response to Block. *Analysis* 71, DOI: 10.1093/analys/anr040
20 Baars, B. (1998) *A Cognitive Theory of Consciousness*, Cambridge University Press
21 Dehaene, S. (2008) Conscious and nonconscious processes. Distinct forms of evidence accumulation? In *Strüngmann Forum Report. Better Than Conscious? Decision Making, the Human Mind, and Implications For Institutions* (Engel, C. and Singer, W., eds), pp. 21–49, MIT Press
22 Dehaene, S. and Changeux, J.P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227

23 Tononi, G. and Koch, C. (2008) The neural correlates of consciousness: an update. *Ann. N. Y. Acad. Sci.* 1124, 239–261

24 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

25 Zeki, S. (2003) The disunity of consciousness. *Trends Cogn. Sci.* 7, 214–218

26 Lamme, V.A. (2006) Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501

27 Rosenthal, D.M. (2008) Consciousness and its function. *Neuropsychologia* 46, 829–840

28 Tononi, G. (2008) Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* 215, 216–242

29 Moutoussis, K. and Zeki, S. (2002) The relationship between cortical activation and perception investigated with invisible stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9527–9532

30 van Gaal, S. *et al.* (2010) Unconscious activation of the prefrontal no-go network. *J. Neurosci.* 30, 4143–4150

31 Cohen, M.X. *et al.* (2009) Unconscious errors enhance prefrontal-occipital oscillatory synchrony. *Front. Hum. Neurosci.* 3, 54

32 van Gaal, S. *et al.* (2008) Frontal cortex mediates unconsciously triggered inhibitory control. *J. Neurosci.* 28, 8053–8062

33 Dienes, Z. (2004) Assumptions of subjective measures of unconscious mental states: higher order thoughts and bias. *J. Conscious. Stud.* 11, 25–45

34 Rees, G. *et al.* (2002) Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3, 261–270

35 Lau, H. 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

36 Rounis, E. *et al.* (2010) Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* 1, 165–175

37 Pleskac, T.J. and Busemeyer, J.R. (2010) Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* 117, 864–901

38 Maniscalco, B. and Lau, H. (2010) Comparing signal detection models of perceptual decision confidence. *J. Vis.* 10, 213

39 Kouider, S. and Dehaene, S. (2007) Levels of processing during non-conscious perception: a critical review of visual masking. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 362, 857–875

40 Hannula, D.E. *et al.* (2005) Imaging implicit perception: promise and pitfalls. *Nat. Rev. Neurosci.* 6, 247–255

41 Lau, H. (2009) Volition and the functions of consciousness. In *The Cognitive Neurosciences IV* (Gazzaniga, M., ed.), pp. 1191–1200, MIT Press

42 Lau, H. and Passingham, R.E. (2007) Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* 27, 5805–5811

43 Pessiglione, M. *et al.* (2007) How the brain translates money into force: a neuroimaging study of subliminal motivation. *Science* 316, 904–906

44 Custers, R. and Aarts, H. (2010) The unconscious will: how the pursuit of goals operates outside of conscious awareness. *Science* 329, 47–50

45 Seitz, A.R. *et al.* (2009) Rewards evoke learning of unconsciously processed visual stimuli in adult humans. *Neuron* 61, 700–707

46 Pessiglione, M. *et al.* (2008) Subliminal instrumental conditioning demonstrated in the human brain. *Neuron* 59, 561–567

47 Hassin, R. and Sklar, A. (2011) Unseen but not unsolved: doing arithmetics non-consciously. *Abstract presented at the 23rd annual convention of the Association of Psychological Science*, Washington, DC

48 Kouider, S. and Dehaene, S. (2009) Subliminal number priming within and across the visual and auditory modalities. *Exp. Psychol.* 56, 418–433

49 Mudrik, L. *et al.* (2011) Integration without awareness: expanding the limits of unconscious processing. *Psychol. Sci.* 22, 764–770

50 Bargh, J.A. and Morsella, E. (2008) The unconscious mind. *Perspect. Psychol. Sci.* 3, 73–79

51 Wegner, D.M. (2003) The mind's best trick: how we experience conscious will. *Trends Cogn. Sci.* 7, 65–69

52 Dijksterhuis, A. *et al.* (2006) On making the right choice: the deliberation-without-attention effect. *Science* 311, 1005–1007

53 Waroquier, L. *et al.* (2010) Is it better to think unconsciously or to trust your first impression?. A reassessment of unconscious thought theory. *Soc. Psychol. Pers. Sci.* 1, 111–118

54 Kouider, S. *et al.* (2007) Cerebral bases of subliminal and supraliminal priming during reading. *Cereb. Cortex* 17, 2019–2029

55 Lau, H. (2008) A higher order Bayesian decision theory of consciousness. *Prog. Brain Res.* 168, 35–48

56 Dehaene, S. *et al.* (2006) Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211

57 de Gardelle, V. *et al.* (2009) Perceptual illusions in brief visual presentations. *Conscious. Cogn.* 18, 569–577

58 Kouider, S. *et al.* (2010) How rich is consciousness? The partial awareness hypothesis. *Trends Cogn. Sci.* 14, 301–307

59 Rahnev, D. *et al.* (2009) Inattention boosts subjective visibility: implications for inattentional and change blindness. *J. Vis.* 9, 157

60 Goldberg, I.I. *et al.* (2006) When the brain loses its self: prefrontal inactivation during sensorimotor processing. *Neuron* 50, 329–339

61 Kepecs, A. *et al.* (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455, 227–231

62 Terrace, H.S. and Son, L.K. (2009) Comparative metacognition. *Curr. Opin. Neurobiol.* 19, 67–74

63 Smith, J.D. (2009) The study of animal metacognition. *Trends Cogn. Sci.* 13, 389–396

64 Kanai, R. *et al.* (2010) Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Conscious. Cogn.* 19, 1045–1057

65 Pollen, D.A. (1995) Cortical areas in visual awareness. *Nature* 377, 293–295

66 Weiskrantz, L. (2009) Is blindsight just degraded normal vision? *Exp. Brain Res.* 192, 413–416

67 Sahraie, A. *et al.* (1997) Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9406–9411

68 Davidson, M. *et al.* (2010) Awareness-related activity in prefrontal and parietal cortices reects more than superior performance capacity: a blindsight case study. *J. Vis.* 10, 897

69 Knable, M.B. and Weinberger, D.R. (1997) Dopamine, the prefrontal cortex and schizophrenia. *J. Psychopharmacol.* 11, 123–131

70 Diederich, N.J. *et al.* (2009) Hallucinations in Parkinson disease. *Nat. Rev. Neurol.* 5, 331–342

71 Vignal, J.P. *et al.* (2000) Localised face processing by the human prefrontal cortex: stimulation-evoked hallucinations of faces. *Cogn. Neuropsychol.* 17, 281–291

72 Duggal, H.S. and Pierri, J.N. (2002) Charles bonnet syndrome: neurobiological insights. *Indian J. Psychiat.* 44, 289–292

73 Ashwin, P.T. and Tsaloumas, M.D. (2007) Complex visual hallucinations (Charles Bonnet syndrome) in the hemianopic visual field following occipital infarction. *J. Neurol. Sci.* 263, 184–186

74 Neander, K. (1998) The division of phenomenal labor: a problem for representational theories of consciousness. *Philos. Perspect.* 12, 411–434

75 Levine, J. (2001) *Purple Haze: The Puzzle of Consciousness*, Oxford University Press

76 Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory and Psychophysics*, Wiley

77 Ratcliff, R. and McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural. Comput.* 20, 873–922

78 Gold, J.I. and Shadlen, M.N. (2007) The neural basis of decision making. *Annu. Rev. Neurosci.* 30, 535–574

79 Beck, J. *et al.* (2007) Probabilistic population codes and the exponential family of distributions. *Prog. Brain Res.* 165, 509–519

80 von Neumann, J. (1958) *The Computer and the Brain*, Yale University Press

81 Libet, B. *et al.* (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106 (Pt 3), 623–642

82 Fried, I. *et al.* (2011) Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron* 69, 548–562

83 Nikolov, S. *et al.* (2010) Probabilistic model of onset detection explains paradoxes in human time perception. *Front. Psychol.* 1, 27

84 McDonald, J.J. *et al.* (2005) Neural basis of auditory-induced shifts in visual time-order perception. *Nat. Neurosci.* 8, 1197–1202

85 Vibell, J. *et al.* (2007) Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *J. Cogn. Neurosci.* 19, 109–120

86 Knight, R.T. and Grabowecky, M. (1995) Escape from linear time: prefrontal cortex and conscious experience. In *The Cognitive Neurosciences* (Gazzaniga, M., ed.), pp. 1357–1371, MIT Press

87 Voytek, B. *et al.* (2010) Dynamic neuroplasticity after human prefrontal cortex damage. *Neuron* 68, 401–408

88 Barceló, F. *et al.* (2000) Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3, 399–403

89 Nakamura, R.K. and Mishkin, M. (1986) Chronic 'blindness' following lesions of nonvisual cortex in the monkey. *Exp. Brain Res.* 63, 173–184

90 Del Cul, A. *et al.* (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132 (Pt 9), 2531–2540

91 Turatto, M. *et al.* (2004) The role of the right dorsolateral prefrontal cortex in visual change awareness. *Neuroreport* 15, 2549–2552

92 Lau, H. (2008) Are we studying consciousness yet? In *Frontiers of Consciousness: Chichele Lectures* (Weiskrantz, L. and Davies, M., eds), pp. 245–258, Oxford University Press

93 Nagal, T. (1974) What is it like to be a bat? *Philos. Rev.* 83, 435–450

94 Heekeren, H.R. *et al.* (2008) The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* 9, 467–479

95 Fleming, S.M. *et al.* (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543

96 Azzopardi, P. and Cowey, A. (1993) Preferential representation of the fovea in the primary visual cortex. *Nature* 361, 719–721

97 Newton, J.R. and Eskew, R.T., Jr (2003) Chromatic detection and discrimination in the periphery: a postreceptoral loss of color sensitivity. *Vis. Neurosci.* 20, 511–521

98 Kim, C.Y. and Blake, R. (2005) Psychophysical magic: rendering the visible 'invisible'. *Trends. Cogn. Sci.* 9, 381–388

99 Davidson, M. *et al*. Different physiological correlates for perceptual decisions and confidence ratings support multi-stage theories. *J. Vis.* 11 (in press)

100 Sperling, G. (1960) The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29

101 Dienes, Z. and Perner, J. (2001) When knowledge is unconscious because of conscious knowledge and vice versa, In *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, 1–4 August, Edinburgh, Scotland. pp. 255–260, Lawrence Erlbaum Associates

102 Brentano, F. (1973/1874) *Psychology from an Empirical Standpoint* (Rancurello, A.C., Terrell, D.B., and McAlister, L.L., trs), Routledge & Kegan Paul

103 Cleeremans, A. (2011) The radical plasticity thesis: how the brain learns to be conscious. *Front. Psychol.* 2, 86

104 Hobson, J.A. (2009) REM sleep and dreaming: towards a theory of protoconsciousness. *Nat. Rev. Neurosci.* 10, 803–813