

## Higher-order awareness, misrepresentation and function

David Rosenthal

*Phil. Trans. R. Soc. B* 2012 **367**, 1424-1438  
doi: 10.1098/rstb.2011.0353

---

### References

[This article cites 65 articles, 18 of which can be accessed free](#)

<http://rstb.royalsocietypublishing.org/content/367/1594/1424.full.html#ref-list-1>

[Article cited in:](#)

<http://rstb.royalsocietypublishing.org/content/367/1594/1424.full.html#related-urls>

### Subject collections

Articles on similar topics can be found in the following collections

[cognition](#) (208 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

*Research*

# Higher-order awareness, misrepresentation and function

David Rosenthal\*

*Program in Philosophy and Concentration in Cognitive Science, Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016-4309 USA*

Conscious mental states are states we are in some way aware of. I compare higher-order theories of consciousness, which explain consciousness by appeal to such higher-order awareness (HOA), and first-order theories, which do not, and I argue that higher-order theories have substantial explanatory advantages. The higher-order nature of our awareness of our conscious states suggests an analogy with the metacognition that figures in the regulation of psychological processes and behaviour. I argue that, although both consciousness and metacognition involve higher-order psychological states, they have little more in common. One thing they do share is the possibility of misrepresentation; just as metacognitive processing can misrepresent one's cognitive states and abilities, so the HOA in virtue of which one's mental states are conscious can, and sometimes does, misdescribe those states. A striking difference between the two, however, has to do with utility for psychological processing. Metacognition has considerable benefit for psychological processing; in contrast, it is unlikely that there is much, if any, utility to mental states' being conscious over and above the utility those states have when they are not conscious.

**Keywords:** consciousness; conscious awareness; metacognition; higher-order theory; misrepresentation; function of consciousness

## 1. INTRODUCTION

Theories of the consciousness of mental states fall into two broad types. One type consists of the so-called higher-order theories, and the other type comprises what have come, in contrast, to be called first-order theories. Higher-order theories all explain what it is for states to be conscious by appeal to an awareness of that state; because it is an awareness of another state, we can call it a higher-order awareness (HOA). No state of which one is not in any way aware is a conscious state. First-order theories, in contrast, deny that a state's being conscious involves any such HOA.

Being aware of a state resembles in some ways the awareness that occurs in metacognitive functioning. So it will be important to see in what ways the HOAs that higher-order theories posit resemble and differ from metacognition, strictly so-called.

Because the HOA that such theories invoke is something distinct from other mental properties of conscious states, it seems possible that a HOA could misrepresent what mental states one is in. And first-order theorists have pressed this as an apparently unintuitive consequence of higher-order theories, undermining their credibility.

Another issue sometimes raised against higher-order theories is whether they allow for a convincing explanation of the function of consciousness, that is,

of the utility that conscious states have specifically in respect of being conscious. This is a second challenge to higher-order theories that is important to evaluate.

In §2, I outline the way higher-order theories seek to explain the consciousness of mental states, and the basic arguments in favour of that kind of theory, and in §3, I consider whether the HOA such theories posit is relevantly similar to metacognitive functioning. In §4, then, I argue that the possibility of misrepresentation by such HOAs is not after all a disadvantage of such theories. Indeed, it is likely that consciousness does actually often misrepresent our thoughts, desires and experiences. And in §5, I argue that a state's being conscious adds little, if any, utility to that which results simply from being in such states when they are not conscious. My argument for this somewhat surprising conclusion does not rely on adopting a higher-order theory of consciousness, but the conclusion fits well with those theories.

## 2. HIGHER-ORDER THEORIES OF CONSCIOUS AWARENESS

A theory of consciousness may serve various explanatory purposes. It may tell us which neural processes subserves a mental state's being conscious. Independent of that, such a theory might tell what it is for a mental state to be conscious, why any states are conscious and perhaps even why particular neural processes do subserves the conscious states they do. However, whatever other explanations a theory of consciousness may provide, it must at a minimum tell us how mental states that are conscious differ from those that are not.

\*[davidrosenthal@nyu.edu](mailto:davidrosenthal@nyu.edu)

One contribution of 13 to a Theme Issue 'Metacognition: computation, neurobiology and function'.

Unless it does that, it must remain unclear that the theory is about consciousness at all [1,2].

### (a) *Higher-order theories*

There is a natural way of understanding how conscious states differ from mental states that are not conscious. No mental state is conscious if the individual that is in that state is in no way aware of it. If somebody thinks, desires or feels something but is wholly unaware of doing so, then that thought, desire or feeling is not a conscious state.

Experimental work on non-conscious perception typically exploits this commonsense observation. Participants sometimes deny seeing a stimulus even when there is evidence, say from priming, that the relevant visual information has affected psychological processing. The effect on subsequent psychological processing, moreover, typically reflects the perceptual discriminations that are characteristic of conscious visual states, e.g. among colours and shapes. We commonly conclude that the visual state occurred but without being conscious. In such cases, a participant's denial of seeing the stimulus reflects not a failure to see, but simply a lack of awareness of seeing. Things are the same outside experimental work. If a person denies wanting something but acts as people typically do when they want that thing, then we see the person as having that desire, though a desire that is not conscious. Novelists and dramatists have described such situations for centuries.

Higher-order theories take this commonsense observation as basic to understanding how conscious states differ from mental states that are not conscious. Because no mental state of which one is wholly unaware is conscious, conscious states are mental states we are in some suitable way aware of. Higher-order theories differ among themselves about just what kind of awareness is required for a mental state to be conscious, but they are agreed that a state's being conscious involves some form of HOA.

When somebody perceives something subliminally, so that the perception is not conscious, there is nonetheless a kind of awareness of the perceived stimulus. It may sound awkward to speak of a non-conscious state that nonetheless makes one aware of something, but we can distinguish the conscious and non-conscious cases in a completely natural way. When one subliminally perceives something, one is aware of that thing but not consciously aware of it; when one consciously perceives the stimulus, one is consciously aware of it.

Traditional theorists [3, p. 115; 4, p. A22/B37] have typically held that the required HOA is perceptual or quasi-perceptual, a view also sometimes championed today [5,6]. Some have argued, however, that the HOA is more likely a thought that one is in the relevant state [1,7,8]. Still others have urged that the HOA is internal to the state one is aware of [9,10], though most see the HOA as distinct from the state it makes one aware of. But all these versions of higher-order theory hold that conscious states differ from mental states that are not conscious in virtue of some HOA [11].

### (b) *First-order theories*

Theories that deny that a state's being conscious consists of one's being aware of it in some suitable way are typically

called first-order theories, though these theories often have little in common beyond that denial. Such theories all arguably face difficulty in explaining how conscious states differ from mental states that are not conscious.

Consider the first-order theory owing to Dretske [12], in which a state is conscious if being in the state results in one's being aware of something else; no awareness of the state is needed. A perception is conscious, then, because it makes one aware of the stimulus. But, Dretske [13] recognizes that this will not do as it stands; subliminally perceiving something makes one aware of it, just not consciously aware of it. To meet this difficulty, Dretske [13] stipulates that a perception is conscious only if the individual can cite the content of the perception as a justifying reason for doing something. However, one cannot cite something one is unaware of; so citing the content of a perception requires one to be aware of that perception. Dretske has ended up invoking higher-order considerations to explain how conscious states differ from mental states that are not conscious.

Other first-order theories seek to explain that difference without any higher-order factors. According to Block's [14] notion of access consciousness, a state is conscious if its content is available for use in reasoning and the rational control of action and speech. In a similar spirit, Baars [15] global-workspace theory and Dehaene *et al.*'s [16,17] neuronal global-workspace theory provide that a state is conscious if it is globally available for psychological processing. Mental states that are not conscious, on these views, lack the relevant availability.

But problems face these explanations. For one thing, it is unclear what such global availability has to do with a state's being conscious. The two seem, at first sight, independent, and they probably are. Visual states towards the periphery of the conscious visual field are presumably not globally available, nor are many conscious but stray passing thoughts. And it is likely that many beliefs and desires that are not conscious nonetheless have widespread, substantial effects on action and psychological processing. So, it is far from clear that global availability can explain how conscious states differ from mental states that are not conscious.

Block [14] also sees conscious qualitative states, which he refers to as phenomenal consciousness, in first-order terms. No HOA figures or is needed on his view for such states to be conscious. Block holds that such states are conscious in virtue of particular sorts of cortical activation; representations in visual cortex, for example, constitute phenomenal consciousness in virtue of suitable cortical activity and connections in visual areas [18,19].

Because neural activity subserves all mental functioning, including the consciousness of mental states, there will be some difference in neural activation for every difference in mental functioning. So knowing that neural activation differs when there is a difference in mental functioning will help us understand such differences in mental functioning.

### (c) *Assessing first- and higher-order theories*

But such appeal to neural activation cannot explain what a difference in mental functioning consists of. For that, we need to describe the mental functioning

in distinctively psychological terms. The appeal to neural activation can help only when we already have an accurate psychological description of the mental functioning. For conscious qualitative states, we need a grasp in psychological terms of what it is for such states to be conscious. And Block [20, §1.3] concedes that, on his first-order view, there may be little or nothing informative to say about that.

These considerations aside, there is in any case reason to doubt that activation in visual areas is all that matters to visual states' being conscious. Lau & Passingham [21] have elegantly isolated cases in which visual performance is matched despite the presence of conscious awareness in some cases but not in others. The only difference functional MRI revealed between cases with and without conscious awareness was activation in mid-dorsolateral prefrontal cortex (PFC). This suggests that the consciousness of perceptual states is not solely due to activation in visual areas, but also to PFC activity. There is substantial additional evidence in support of that conclusion [22].

The hypothesis that PFC activation is needed for conscious awareness fits well with the higher-order theories of consciousness. Activation in visual areas provides both visual content and visual qualitative character, but by itself does not result in conscious visual awareness. Only when there is also relevant PFC activation do those visual states become conscious. Evidently, PFC activation subserves the relevant awareness of the visual states, which on their own occur without being conscious. Discoveries about neural activation do not by themselves explain how conscious states differ from mental states that are not conscious, but they sustain a higher-order explanation cast in distinctively psychological terms.

When visual content is transmitted from visual areas to PFC, the relevant visual states are conscious [16,17]. But that finding cannot by itself decide between the neuronal global-workspace theory and a higher-order theory of consciousness. Neural activation in PFC might be relevant because it makes the representational content of the visual states globally available, but it might instead subserves an awareness of the relevant perceptual states.

PFC subserves many distinct psychological processes. So PFC activation when states are conscious but not when they fail to be conscious cannot by itself decide between a higher-order theory and the neuronal global-workspace theory. But further investigation is needed to determine whether the PFC activation that accompanies states' being conscious serves to make their content globally available or instead subserves HOAs of those states. Making a state's content globally available is a distinct psychological process from producing an awareness of that state; the neuronal global-workspace theory posits that the first type of process is responsible for a state's being conscious, whereas higher-order theories explain that by appeal to the second type of process.

In any case, we have seen that there is reason, independent of any neural findings, to prefer the higher-order theory. Many conscious states are doubtless globally accessible to various cortical systems, and many non-conscious states are not thus accessible, as

work by Dehaene and co-workers [16,17] has made evident. But because counterexamples exist to both generalizations, the appeal to global availability cannot explain why some states are conscious and others not. Nor, as already noted, is that surprising; even if most conscious states are globally available and most non-conscious states are not, it is unclear what a state's being globally accessible has specifically to do with its being conscious. So, even though PFC activation by itself lends support both to higher-order and to global-workspace theories, only the appeal to HOAs underwrites a satisfactory explanation of why some states are conscious and others not.

In our everyday dealings, we have little, if any, concern with mental states that are not conscious. Circumstances must become somewhat special before we ask whether somebody has a thought, desire, perception or sensation that is not conscious. Such circumstances apart, when somebody denies being in such a state, we take that denial at face value. We may sometimes have reason to override such denials, perhaps more often with desires and emotions than with bodily sensations, but we rarely have reason with any kind of mental state to doubt others' sincere denials.

This everyday practice of crediting what others say about what mental states they are in can encourage seeing conscious states as the norm, and non-conscious states, such as subliminal perceptions, as at best degenerate cases of such states. Crediting others' sincere reports about their mental states leads to our discounting the possibility that they are in mental states they are unaware of, and so that any of their mental states fail to be conscious. We rely on their first-person awareness of mental states to determine what mental states they are in, tout court, thereby collapsing the distinction between mental states and conscious mental states.

All this may seem to support the adoption of a first-order theory of consciousness. If, as Block [23] suggests, the default for mental states is that they are conscious, then perhaps we should not seek an explanation cast in distinctively psychological terms of how conscious states differ from mental states that are not conscious. And then there may seem no need to invoke awareness of some mental states to distinguish the conscious cases from those that are not conscious.

This explanatory retreat would make it appealing to find some purely neuronal explanation of what it is for a state to be conscious, or an explanation that appeals to global availability. We would no longer be concerned to explain how conscious states differ from mental states that are not conscious, but only to give some neuronal condition to occur whenever conscious states occur. In addition, if we took being conscious to be the default for mental states, it might seem tempting to regard any HOA that does figure in mental states' being conscious as somehow intrinsic or internal to mental states [10], a theory Block has called same order [19,24].

But we should not permit our everyday practice of taking people's sincere views about what mental states they are in to influence theorizing about consciousness. We have ample evidence of individuals' being in mental states they are unaware of, and their views and remarks about what states they are in can

at best reflect only their awareness of what states they are in. Verbal reports indicate reliably what states an individual is aware of, but not what states occur without the individual's being aware of them [25,26]. We can understand the difference between conscious and non-conscious mental states only by appeal to whether there is a suitable HOA.

Nor is there good reason to see such HOAs as intrinsic or internal to the states they make one aware of. For one thing, findings by Libet *et al.* [27] and Haggard [28] show that the neural occurrence that leads to a particular action occurs in advance of any awareness of a volition to perform that action. We can best regard the antecedent neural event as the volition to perform an action, initially occurring without being conscious and only subsequently becoming conscious [29]. It is difficult to square these findings with the hypothesis that the HOA of a volition or other mental state is intrinsic to that state.

More generally, a theory on which HOAs are intrinsic to the states they make one aware of must provide an independent reason to individual mental states that yield this result. The default assumption would be that a HOA is distinct from the first-order state because they are about different things. We can therefore best see the hypothesis that HOAs are intrinsic to the states they are about as an attempt to split the difference between first- and higher-order theories. The hypothesis concedes that no state is conscious without some HOA, but nonetheless joins first-order theories in positing nothing beyond the conscious state itself. But there is no theoretical advantage to this attempted marriage of first- with a higher-order theory; the explanatory work is all done by the HOA.

### 3. METACOGNITION AND HIGHER-ORDER AWARENESS

The term 'metacognition' covers a broad range of phenomena in which individuals have some knowledge or sense of their own cognitive functioning. That suggests that the awareness we have of mental states when those states are conscious may be a type of metacognition. Many first-order mental states, perhaps all of them, represent the world as being one way or another, and are in that way cognitive. So being aware of oneself as being in some such state looks at first sight like a form of metacognition. The states posited by higher-order theorists are higher-order in virtue of their higher-order intentional content—that is, content that is about another mental state. And the judgements that figure in metacognition also have higher-order content, that is, content about first-order cognitive states.

#### (a) *Differences between metacognition and higher-order awareness*

However, the phenomena typically classified as metacognitive differ in crucial ways from these HOAs. Indeed, there is reason to believe that standard types of metacognition occur without any conscious awareness at all [30].

Standard types of metacognition have to do with whether something currently being learned or having

previously been learned will be readily recalled in the future. Nelson & Narens [31] describe judgements about information that cannot currently be recalled, but likely will be in the future, as feeling-of-knowing (FOK) judgements; they describe judgements that information currently available will continue to be as judgements of learning (JOL). Others distinguish the two in slightly different ways. In the tip-of-the-tongue (TOT) phenomenon, one may feel one knows, e.g. what George Eliot's real name was, though one cannot currently recall it; some see TOT as a type of FOK judgement, and others not [32].

Each of these types of metacognitive judgement concerns conscious availability of some first-order informational content. But those judgements differ strikingly from the HOAs that result in one's being aware of a mental state. In TOT and FOK, the first-order informational state is not currently conscious, and the metacognitive judgements concern only future conscious availability of the relevant information. In JOL, though the first-order informational state is currently conscious, the metacognitive judgement again pertains only to future conscious availability. And although all these judgements are about first-order informational states, they are in each case predictions about future recall of the relevant information. Like HOAs, metacognitive judgements have higher-order content because they are about other mental states, but unlike HOAs, they do not operate to make one aware that one is in the state they are about.

These metacognitive judgements differ in another crucial way from the HOAs posited in higher-order theories. We are seldom aware of any such HOAs. Many mental states are conscious, and when they are, we are aware of those states. However, we are rarely also aware of any higher-order states directed upon them. Higher-order theories predict this; no HOA would itself be conscious unless, in addition to the HOA, there were a third-order awareness that made one aware of that second-order awareness. We can expect that such third-order awarenesses are rare. When one introspects some conscious state that one is in, one is then aware of focusing attentively on that state; so one is aware of one's awareness of the state. A third-order state occurs in introspective awareness of a conscious state, but not otherwise. It is a disadvantage of the view that HOAs are intrinsic to the states they make one aware of that they make the wrong prediction about this, holding that we are aware of all HOAs [10].

The metacognitive judgements that occur in FOK, TOT and JOL, in contrast, are all conscious judgements. In experimental work on metacognition, participants are plainly aware of having those judgements [32,33]. But these judgements are not HOAs because their content does not describe one in terms of current mental states. Rather, they are judgements about likely future recall of information, that is, about what states one may come to be in. So although the metacognitive judgements are conscious, it does not involve third-order, introspective awareness of a current state.

The contrast just drawn between HOAs and metacognitive judgements may strike one as paradoxical.

HOAs are seldom themselves conscious states, yet they result in the relevant first-order states' being conscious; metacognitive judgements are often if not always conscious, and yet typically do not result in first-order states' being conscious. How can that be? Moreover, how can HOAs that are not themselves conscious make the first-order states they are about conscious?

On higher-order theories, first-order states do not inherit the property of being conscious from higher-order states. On such theories, the property of a state's being conscious consists of one's being aware of oneself as being in that state, and the higher-order states constitute those awarenesses. The HOA does not pass along the property of being conscious to the first-order state; it simply serves to make one aware of that state in the right way, and that is what the state's being conscious consists of.

In contrast, many metacognitive judgements do not make one aware of oneself as being in some first-order state. Rather, the metacognition states in JOL and FOK are judgements about what knowledge one has acquired and what knowledge one will retain and be able to produce at will. Although metacognitive judgements, such as HOAs, are about first-order states, many metacognitive judgements do not have the content that one is now in a particular state.

#### (b) *Tip-of-the-tongue phenomena*

Some care is needed in understanding how consciousness figures in TOT. Although one is unable in such cases to recall the specified information, e.g. George Eliot's real name, one has a sense, often quite compelling, that one will at some time readily recall it. So, one regards oneself as having the information despite one's current inability to recall it.

Because one is aware of oneself as having the information, one regards oneself as having the belief, say, that George Eliot's real name is Mary Anne Evans even though one cannot just now get at the content of that belief. Indeed, the sense that one has such a belief is typically quite compelling. So, one is aware of oneself as being in the relevant mental state. Is this a counterexample to higher-order theories, on which being suitably aware of being in a mental state suffices for that state to be conscious? If not, why is the state that carries the information that George Eliot's real name is Mary Anne Evans not conscious in the TOT situation?

In TOT, one is aware of a mental state that, as it happens, carries that information; one is aware of being in a state whose informational content would say what George Eliot's real name is. But one is not aware of that state in respect of that specific information. One is aware of the state only in respect of one aspect of its intentional content, namely that it has intentional content that would answer the question of what George Eliot's real name is. It is the awareness of the state in respect of only an aspect of its intentional content that is responsible for the notorious subjective oddness of the TOT situation.

Does that awareness result in the state's being conscious? It may not seem obvious what to say. The awareness we have of our conscious states does not always reveal all their representational character; indeed, HOAs seldom capture every aspect of the

first-order state. Still, it seems misleading, at best, to say that the TOT feeling results in the informational state's being conscious. That is because the HOA leaves out too much of the state's content; one would regard the state as conscious only if one were aware of the state in respect of that aspect of the content that is of current interest [34].

Metacognition is not confined to TOT, FOK and JOL. Change blindness consists of a failure to detect many visual changes in a scene, often changes that are relatively salient. However, people are also typically unaware of such failures to detect such changes. Levin *et al.* [35] regard this as a metacognitive failure, which they refer to as change blindness blindness. It is a metacognitive failure in reflecting a judgement that people generally detect any significant change that occurs in a visual scene. And many authors think of metacognition in similarly broad terms, as even encompassing HOAs of one's conscious states [36]. The present discussion will confine itself, however, to types of metacognition along the lines of JOL, FOK and TOT.

#### (c) *Metacognition and conscious awareness*

Not every way of being aware of a mental state results in that state's being conscious. As many have noted [3, p. 335; 37, pp. 77, 171], the awareness must be subjectively unmediated; when one's awareness of being in a state seems subjectively mediated, the state is not conscious.

Suppose I behave happily or angrily or as though I would like to have a particular thing, but I sincerely deny that I am happy or angry or that I desire that thing. We have reason to conclude that I am in the relevant state, though the state is not conscious. Now you tell me that I am behaving in a way that suggests that I am actually in the state and, having confidence in your judgements about such things, I believe you. I become aware of being in the state, but as long as I am thus aware only because I believe what you tell me, that awareness does not result in the state's being conscious.

But it could happen instead that your telling me triggers an independent awareness of being in the relevant state. Although I would not have been aware of the state without your remark, I am now aware of the state independently of taking your word for it; I would be aware of it even if I came to see your judgement as in some way ill-founded. My awareness is subjectively unmediated, and the state is conscious. We can explain this by saying that the awareness does not seem subjectively to depend on any inference or self-observation; if asked, I would not cite such factors as a basis for my awareness that I am in the state. One does not, contrary to Block's [38] suggestion, also need an additional HOA to the effect that there is no inference.

Traditional theorists [3, pp. 335, 592; 9; 37, p. 113] maintained that one's awareness of a state must actually be direct, but that overshoots; the awareness need only seem to be direct. So the awareness might actually depend on some inference or self-observation so long as it does not subjectively seem that way to the individual.

Applying this condition requires care. Suppose a therapist tells one that one is in a deeply repressed state, and one believes that. And suppose that one then forgets who told one, and hence how one learned

of the repressed state, but one still believes one is in it. Block [38, p. 446] argues that this should result in its seeming to one that one's awareness is independent of any inference or self-observation. But simply forgetting the source would by itself not result in one's awareness of the state seeming subjectively not to depend on any inference or self-observation. Unless something else occurred, it would still seem that one's awareness rested on some such mediating factors, despite one's inability to recall which. Such a case would, in that way, resemble TOT; one would have a compelling sense that one had learned from some source that one was in the state, but could not recall what the source was. Only if one became aware of the state in a way that did not seem subjectively to rely on some such source or other would the state come to be conscious.

Like the HOAs that result in mental states being conscious, metacognitive judgements seem unmediated or direct; it does not seem subjectively that they depend on any inference or self-observation. The crucial difference between HOAs and metacognitive judgements lies in their content. The metacognitive judgements considered earlier are judgements about what information one is likely to recall in the future; HOAs all have the content that one is currently in the state in question.

Although metacognitive judgements are typically conscious [39], Kentridge & Heywood [40] have noted that metacognitive processing can occur without conscious awareness at all. G.Y., an individual with blindsight, was able to adapt his visual processing in response to a visual cue presented within the blind field. As Kentridge & Heywood note, such adaptation is typically taken to reflect monitoring of one's psychological processing, and hence is seen as metacognitive in nature. Although striking, this finding should occasion no surprise, because metacognitive judgements are distinct from the type of awareness that results in mental states' being conscious.

Some who deny that a state's being conscious consists of being aware of that state in a suitable way have stigmatized any such awareness as metacognitive awareness, irrelevant to the state's being conscious. Thus, Seth [41] writes that 'sensory content need not be overlain by metacognitive content in order to be conscious' (p. 981). And Block [14] sees HOAs as occurring only when a state is reflectively or introspectively conscious, not when it is conscious in a non-introspective way. This evidently fails to take into account that the awareness that higher-order theories posit is seldom itself conscious. When a HOA is conscious, the state is introspectively conscious, and the awareness of it does seem something like standard types of conscious metacognitive judgement. A state of which one is wholly unaware is not in any intuitive way a conscious state, but the needed awareness typically falls well short of metacognitive processing as that is usually understood.

#### **(d) Reports, confidence measures and post-decision wagering**

Reliance on an individual's report of being in a state or not being in it to determine whether the state is conscious [25,26,42] is known as a subjective measure

[43]. Such measures reflect a higher-order theory of what it is for mental states to be conscious, because a report that one is in a state expresses an awareness of being in that state, and similarly for a report that one is not.

An alternative measure, also subjective in nature, is to have participants rate the degree of confidence they have about a particular stimulus. Instead of relying on explicit reports about whether participants are in particular states, one can ask them to rate how confident they are that a perceptual decision about a stimulus is accurate. High confidence suggests that participants have conscious awareness of the stimulus. Low confidence, in contrast, suggests that participants take themselves just to be guessing; a highly accurate perceptual decision with low confidence suggests that a participant is perceptually detecting the stimulus, but not consciously. Such confidence ratings may therefore reveal whether a HOA is present, and they differ from cases of metacognition such as JOL and FOK, because such ratings are described in terms of accuracy of a current perceptual judgement, not in terms of whether one has or will have command over some particular information.

Persaud *et al.* [44] have developed a test for mental states' being conscious that arguably improves on such confidence ratings. Instead of explicitly rating their degree of confidence, participants are asked to place a wager on a perceptual decision they have made, say, about whether a stimulus is present or a string of letters exhibits a particular pattern. Wagering presumably reflects degree of confidence, but motivates subjects to act on that degree of confidence and so may avoid methodological issues about confidence ratings [45].

Schurger & Sher [46] and Dienes & Seth [47] independently pointed to the difficulty that loss aversion in some subjects [48] creates for the wagering test, and have suggested ways to adjust for that. But Dienes & Seth have shown that even when wagering is adjusted for loss aversion, it is no more sensitive than traditional confidence ratings as an indicator of conscious awareness. See also Overgaard & Sandberg [36] for a useful review of the wagering method for determining confidence ratings.

There is, moreover, a question about whether the non-conscious information that influences conscious perceptual decisions might also affect conscious wagering behaviour. Presumably, it does not typically do so, because post-decision wagering results conform well to traditional subjective reports [44,47], and people may typically base wagers only on information they are consciously aware of. But unconsciously seen stimuli do affect inhibitory mechanisms [49], and so might also affect wagering. That would influence traditional confidence ratings as well as post-decision wagering, leaving subjective reports as the standard against which other measures must be calibrated.

#### **4. HIGHER-ORDER MISREPRESENTATION**

Metacognitive judgements are by no means always accurate. JOL may misgauge how much one will

recall [50], and feelings of knowing can misrepresent how much information one has command over [51]. Indeed, various techniques have been proposed for enhancing the accuracy of metacognitive judgements [52]. And change blindness illustrates a routine, persistent error in the metacognitive expectation that we readily detect visual changes [53,35].

Metacognitive error is hardly surprising. Although metacognitive judgements are often sensitive to actual cognitive abilities and achievements, such judgements are distinct from the cognitive states they are about. So, there will always be room for metacognitive judgements to be wrong about those cognitive states.

#### (a) *Misrepresentation by conscious awareness*

This suggests a parallel question about the consciousness of mental states. On higher-order theories, a state is conscious in virtue of one's being aware of that state, and that awareness is something distinct from the mental properties it makes one aware of. This is clear on standard higher-order theories, which hold that the HOA and the state it makes one aware of are distinct states. But even on higher-order theories that maintain that the HOA is intrinsic to the target state [9,10], the HOA is distinct from those mental properties it makes one aware of and in respect of which we taxonomize the conscious state. Because the HOA is distinct from the state one is aware of, or at least distinct from the mental properties in respect of which one is aware of it, such theories allow for possible misrepresentation by consciousness. In allowing for such misrepresentation, consciousness resembles metacognition.

Consciousness is the way our mental lives appear to us; it is mental appearance. And because appearances are not always accurate, there should be no surprise that consciousness can misrepresent the mental states that actually occur in us. Consciousness may make it appear subjectively that one is in a mental state somewhat different from the state one is actually in.

But there is a traditional tendency to think that, unlike other appearances, the appearances of consciousness cannot diverge from reality, that in respect of the appearances of consciousness the mind is transparent to itself [3, pp. 364, 592; 4, p. B132; 9; 37, pp. 77, 171]. If so, consciousness could never misrepresent the mental reality it makes us aware of; when it comes to consciousness, there is no distinction between appearance and reality [54]. Consciousness would not merely be the way our mental lives appear to us; it would actually constitute that mental reality. Anything we think we know about the mind would then have to be tested against the deliverances of consciousness. If appearance and reality coincide for consciousness and mind, then misrepresentation by consciousness is plainly impossible. Because higher-order theories accommodate that possibility, such theories could not then be correct.

It is worth stressing that the apparent difficulty for higher-order theories is not about whether consciousness does sometimes misrepresent what mental states individuals are in, though there is reason to think it actually does. But the alleged difficulty concerns whether such misrepresentation is even possible. A number of

theorists have pursued this criticism, urging that the mere possibility of misrepresentation leads to incoherent or absurd consequences [24,38,55–57]. Any theory that allows consciousness to misrepresent mental reality must, they maintain, be mistaken about the very nature of consciousness and mind.

#### (b) *Kinds of higher-order misrepresentation*

The idea that misrepresentation is absurd or incoherent recalls the traditional idea that first-person access to mental states is infallible. Only if first-person access is infallible would misrepresentation by consciousness be impossible. But it is plain that first-person access is not infallible. Mental states sometimes occur without one's being in any way aware of them, as in subliminal perception and other cases. Consciousness in these cases erroneously rules that no such states occur. There are mental states we can detect even when consciousness leads subjects to deny their occurrence.

One can always seek to defend a theory by redescribing the data. So one might deny that subliminal states are mental at all, perhaps describing them as subpersonal states, mere 'events of content fixation', as Dennett [58] urges. Or one could urge, with Block [14,18,19], that even some perceptual states that one is wholly unaware of and sincerely denies the occurrence of are conscious in a special way that does not require any such awareness. But states that occur in subliminal perception function in many or even most of the ways that conscious perceptual states function [59]. So denying that subliminal states are qualitative or even mental at all is simply redescribing commonsense phenomena to save a theoretical preconception. And if first-person access is not infallible, there can be nothing incoherent or absurd about the possibility of misrepresentation by consciousness.

A finding by Breitmeyer *et al.* [60] (see also Breitmeyer *et al.* [61]) may seem to provide a substantive reason to question whether genuinely qualitative states do occur in subliminal perception. Breitmeyer *et al.* found that in metacontrast masking of unsaturated blue, green and white stimuli, the priming effects of the white stimulus resemble those of the green stimulus more than the blue. Because the white stimulus had greater contributions of green wavelengths than of blue, Breitmeyer *et al.* concluded that priming effects in these masked cases reflect earlier visual processing in area V1 that is mainly responsive to wavelength properties. This effect was not found when stimuli were unmasked and consciously visible. Green stimuli in those cases did not prime like white stimuli, presumably reflecting cortical activity in higher visual areas specialized for colour. Breitmeyer *et al.* described the masked effect as wavelength-dependent, in contrast with what they described as the percept-dependent behaviour of the unmasked primes.

One might take this to show that the mental qualities characteristic of conscious vision do not occur in subliminal processing. But the findings do not support that conclusion. Processing in the higher visual areas specialized for colour may typically be conscious, but these findings do not show that it always is. Nor is it obvious that such processing never fails to be

conscious. So if the masked cases do not exhibit mental qualities, that need not be because those cases fail to be conscious; it may instead be because they do not involve processing beyond area V1. It may even be that mental qualities do figure in the masked cases, but that the lack of higher visual processing results in those qualities being taxonomized in the more coarse-grained ways reflected in the findings by Breitmeyer *et al.* More would be needed to show that mental qualities do not occur at all in subliminal vision.

Consciousness may misrepresent mental reality, then, by failing to reveal non-conscious mental states of which one is altogether unaware. But how about states that are conscious? Can consciousness fail to represent those states accurately? Not only is that possible; it occurs routinely. When we consciously take in a visual scene, we see the colours that things have, but we are seldom aware of those colours in respect of their exact shades. We consciously see things as light blue or dark red or bright green, but we are not consciously aware of the exact shades we would discern if we attended to the objects and their colours.

One might question whether being aware of specific shades of visual sensations in merely generic terms is properly described as a case of misrepresentation; we do not generally describe cases of incomplete representation as misrepresentation. But in the cases under consideration, first-order states simply have no generic mental properties of the sort that HOAs represent those first-order states as having; the mental properties of first-order states are all specific. So HOAs do sometimes represent the first-order states inaccurately.

There are, in any case, more dramatic cases that we would plainly describe as misrepresentation. Grimes [62] used eye trackers to switch displays during participants' saccades, when no retinal signal reaches visual cortex. Participants often were not consciously aware of changes in salient visual features, e.g. in 18 per cent of cases, a dramatic change between red and green central to a display. Because retinal input to visual cortex resumes after saccades, first-order states in visual cortex presumably did change in ways that reflected the change in display, despite participants' reporting no awareness of such change. Participants' HOAs of their first-order visual states here plainly misrepresent those states.

But the visual states themselves, independent of conscious awareness, must reflect the more finely differentiated shades, because those shades are readily available whenever we attend to them. The same is true of the mental qualities we are aware of inattentively in other perceptual modalities. Mental qualities occur in more finely differentiated ways than consciousness typically reveals. This is one way in which consciousness misrepresents mental reality, though the misrepresentation is innocuous and is corrected at will by attending. Consciousness misrepresents in connection with other types of mental state as well; Desmurget and co-workers [63,64] have found evidence of awareness of intentions without intentions, and conversely; here, consciousness represents volitional states inaccurately.

It is worth noting a possible difference between these cases and those reported by Wegner [65], in which participants are aware of themselves as doing something and thereby causing something to happen even though

those effects are due to independent factors. These participants are aware of themselves as being in relevant volitional states and also aware of themselves as causing the target effects. But those two cases of awareness are distinct and can occur independently. Participants might have the volitions they are aware of having even though those volitions do not figure in causing the target effects. Additional evidence would be needed to show that they also lack those volitions.

### **(c) Attention, conscious awareness and misrepresentation**

Attending cannot by itself always yield a more finely differentiated awareness of mental qualities. We are consciously aware of shades of colour in more finely differentiated ways when we have two or more very close shades to compare at one time than when those shades are presented in succession [66,67]. This also holds for conscious discrimination of auditory pitch [68]. We have no reason to hold that visual and auditory cortex represents shades and pitches in less fine-grained ways when they occur successively rather than together. So when stimuli do occur in succession, consciousness fails to reveal fine-grained differences that actually occur among the relevant mental qualities. And this effect is altogether independent of attention.

Attention enhances the way we are aware of our perceptual experiences. But one might urge that some attention, perhaps quite low, is needed for a HOA to form in the first place. This is unlikely. For one thing, there is now ample evidence of a double dissociation between attention and mental states' being conscious [69–73]; attention occurs in connection with states that are not conscious, and is absent with many states that are conscious. And it is in any case subjectively implausible that attention is needed for states to be conscious; many peripheral visual states are conscious but seemingly unattended. One could urge that there is always some attention, however slight; when a state is conscious, but without independent evidence of attention, it is unclear what that claim amounts to.

If attention is not responsible for the generation of HOAs, what is? I argue in §5 that HOAs contribute little or no utility over and above the utility of the first-order states themselves; so we cannot appeal to evolutionary selection pressures to explain why HOAs occur. This issue would go beyond the scope of the present study, though I have addressed it elsewhere [1, ch. 7, §6, ch. 10, §5].

### **(d) Other doubts about higher-order misrepresentation**

Block [38] has urged in reply to me [74] that one cannot be consciously aware of a stimulus in respect of a less finely differentiated shade than is represented in visual cortex. But he misconstrues the claim, arguing just that there cannot 'be an experience of red but not of any shade of red' [38, p. 445]. The claim is not that a conscious experience of red might not represent the shade as red at all, but that it might represent it in a way that is indeterminate with respect to a range of specific shades of red. One is aware of the experience as an experience of red, but not,

e.g. as more like maroon than crimson. This kind of conscious awareness is not only possible; it is routine.

It is important to note a possible ambiguity of the term ‘experience’. By ‘experience’, one might mean a sensation or perception of red, independent of whether and if so how that state is conscious. Any such state will exhibit a mental quality of a specific shade of red. But things are different if by ‘experience’ one means instead a conscious sensation or perception. Even though the sensation or perception itself, independent of how it is conscious, will exhibit a mental quality of a specific shade, the way we are aware of the state will often fail to reflect that specific shade. So if by ‘experience’ we mean the qualitative state as one is subjectively aware of it, an experience of red may well simply be of a generic shade.

Failing to capture specific qualitative character is not the only way that consciousness misrepresents our mental states. People sometimes confabulate the thoughts or desires as having led to choices, despite evidence that those thoughts and desires could not have been operative. They are, in these cases, aware of themselves as having thoughts and desires that they do not actually have [75–77]. Moreover, expectations distort the way we are subjectively aware of our qualitative experiences, as when one drinks apple juice expecting the taste of iced tea. Expectations can affect even the subjective experience of pain [78]. In such cases, one’s subjective awareness misrepresents the qualitative state one is actually in.

Why, in the face of such commonplace examples as iced tea and apple juice, would anybody doubt that consciousness sometimes misrepresents actual mental occurrences? Consciousness is the way those mental occurrences subjectively appear to one; why think it cannot get things wrong?

Perhaps, if the subjective awareness were intrinsic not simply to the state, but to the represented mental properties themselves, that would prevent consciousness from misrepresenting. There is a compelling sense that nothing mediates between the painfulness of a conscious pain and one’s subjective awareness of it; the painfulness and the awareness seem indistinguishable. And if they are not only indistinguishable but the same, perhaps that awareness cannot misrepresent the state in respect of such painfulness.

Its seeming that mental qualities are indistinguishable subjectively from awareness of those qualities appears to point toward a first-order theory of consciousness, on which qualitative states are conscious independent of any awareness of them. And when mental states are conscious, the mental properties in virtue of which they are conscious can seem subjectively inseparable from their being conscious.

But such apparent inseparability does not show that consciousness is built into those other mental properties, independent of one’s awareness of those properties. When a state is conscious, one is aware of the state in respect of various individuating mental properties, such as its qualitative character. But one is seldom also aware of being aware of those other mental properties. So the state’s being conscious typically seems subjectively inseparable from the other mental properties in virtue of which one is aware of

it. And the occurrence of states that have those mental properties without being conscious, which pose a challenge for first-order theories, underscores the independence of a state’s being conscious from its other mental properties.

There is independent reason to reject a first-order theory on which a state’s being conscious is inseparable from the state’s other mental properties. Whatever our subjective impressions about conscious states, we can give full accounts of those other mental properties without appeal to consciousness. This is widely recognized for cognitive and volitional states. Despite many disagreements about how best to explain intentionality, virtually every theory with any level of broad support seeks to explain intentionality independent of the relevant states’ being conscious.

Things may seem less clear in the case of qualitative states, because many theorists see mental qualities as inexplicable apart from the way those qualities present themselves to conscious awareness. But as these theorists often acknowledge, this approach permits virtually nothing informative to be said about mental qualities [20,79]. So there is reason not to rely on consciousness in giving an account of mental qualities [80].

And there is an inviting alternative. Instead of understanding mental qualities in terms of the way they present themselves to consciousness, we can understand them by appeal to the roles they play in perception. We discriminate among various ranges of perceptible properties of stimuli, and we are able to do this because of the mental qualities special to each perceptual modality. For every perceptual discrimination between two perceptible properties, there must be a difference between corresponding mental qualities. So when we construct the quality space of perceptible properties discriminable by a particular modality, we have an account of the mental qualities that enable those discriminations [1,80]. Because perceptual discrimination can occur subliminally, this account again makes no appeal to qualitative states’ being conscious.

When mental states are conscious, we individuate them not by appeal to their being conscious, because all conscious states have that in common, but rather by appeal to their other mental properties. The mental properties in terms of which we individuate conscious states are the intentional and qualitative properties we are subjectively aware of those states as having. The availability of satisfactory theoretical accounts of those properties independent of consciousness trumps whatever subjective appearance there is that being conscious is intrinsic to those properties. And that leaves us with no reason to doubt that consciousness does sometimes misrepresent the states it makes us aware of.

### (e) *Neuronal global-workspace theory*

As noted in §2, global-workspace theory is unusual among first-order theories in providing a distinction between mental states that are conscious and those that are not. So it should not be surprising that, unlike other first-order theories, it does not preclude misrepresentation by consciousness. (I am grateful to Hakwan Lau for raising this point.) A state is conscious on global-workspace theory if its content comes to be present in the

global workspace and thereby available to many cortical systems. So consciousness in effect represents the state as having whatever content is available to those systems. But the content that reaches the global workspace might sometimes differ from the content of the original state. And consciousness would then, on the theory, misrepresent the state that is conscious.

One might question why, given various similarities between higher-order theories and the neuronal global-workspace theory, we should not see them as theories of the same type. As noted in §2, both theories receive support from PFC activation when states are conscious and relative absence of such activation when states are not. Both theories allow, moreover, for a type of misrepresentation by consciousness.

But there are pivotal differences. Most importantly, the neuronal global-workspace theory sees a state's being conscious as consisting of the global availability of its content, whereas higher-order theories see a state's being conscious as consisting of one's being aware of oneself as being in that state. These are strikingly different ways to understand the property of a state's being conscious.

The two models of what a state's being conscious consists, moreover, result in distinct ways in which consciousness can misrepresent. On higher-order theories, one is aware of oneself as being in a state in some ways distinct from the state one is actually in. On the neuronal global-workspace theory, in contrast, the misrepresentation is due rather to a difference between the content that the global workspace makes available and the content of the original state itself.

The neuronal global workspace is posited to make content available to many systems, and these may well include whatever system subserves HOA and the consequent ability to report one's own conscious states. So there is that much overlap between higher-order theories and the neuronal global-workspace theory. But it is the HOA that matters for a state's being conscious, according to higher-order theories, not availability to other cortical systems. And experimental testing for most theories must rely on subjects' verbal reports or the equivalent to determine whether a state is conscious.

#### (f) *Misrepresentation and what it is like*

On a higher-order theory, a state is conscious in virtue of one's being aware of oneself as being in that state. On a standard construal of the phrase, 'what it is like' [54], what it is like for one to be in a state is a matter of how it is for one to be in that state, that is, how one is subjectively aware of oneself as being in the state. So the HOA one has of being in a state will determine what it is like for one to be in it.

On some higher-order theories, my own included, that HOA consists of a thought to the effect that one is in the relevant state. And one might question whether such thoughts could be responsible for there being something it's like for one [38, p. 444]. But there is reason to think that it can [1]. Learning new words for one's qualitative experiences can sometimes by itself result in coming to be aware of one's experiences in more fine-grained ways, as with learning terms for tastes of wines or for musical instruments one could not previously distinguish.

Learning new words involves learning how to apply the concepts that the new words express. And learning a new concept enables one to have a new range of thoughts that employ that concept. The new words and concepts in these cases pertain to one's qualitative experiences; so learning the new words enables one to have new thoughts about one's own qualitative experiences, thoughts that draw more fine-grained distinctions among them.

There is no other effect that learning new words might have that would enable one to be subjectively aware of those more fine-grained qualitative distinctions. So the best explanation for this effect of learning new words that draw more fine-grained qualitative distinctions is that the new concepts enable one to have more fine-grained higher-order thoughts about the relevant qualitative states. This demonstrates that differences in the intentional content of higher-order thoughts makes a corresponding difference to the way we are subjectively aware of qualitative states. It therefore lends credibility to the claim that higher-order thoughts are indeed responsible for there being something it's like for one to be in conscious qualitative states in the first place.

This effect is seldom evident in adulthood, because it is rare past young childhood that we learn new terms for our qualitative states, and hence new ways to be aware of those states. So it is useful to focus on such unusual cases as coming to have more fine-grained conscious experiences of wines or musical instruments.

It may seem that higher-order-thought theory makes mistaken predictions about what conscious experiences are possible. As Block [38, p. 445] urges, because nothing rules out having a thought that one has an experience of red and green fully occupying the same region at once, perhaps higher-order-thought theory predicts that such an experience is possible.

Surprisingly, something like that conscious experience seems actually to occur; in a striking study, Billock *et al.* [81] found that when multi-coloured, equi-luminant images are stabilized on the retina, some subjects report seeing reddish greens or bluish yellows, presumably owing to filling-in mechanisms. But higher-order theories are in any case not committed to such predictions. Higher-order thoughts must have an assertoric mental attitude, because doubting and wondering do not by themselves result in one's being aware of anything. And one cannot have any assertoric thought one chooses. It is doubtful that one can assertorically think that  $2 + 3 = 57$ , or even that one can have an assertoric thought that it is raining as one gazes out on a clear, cloudless sky. So it is likely that what mental states we can be in constrain in various ways what assertoric higher-order thoughts we can have.

## 5. FUNCTION, CONTENT AND PSYCHOLOGICAL ASCENT

Metacognitive judgements have often considerable utility; it is plainly useful for a person or other creature to have some sense of how much has been learned (JOL) and how much will be able to be recalled (FOK). Moreover, metacognitive regulation of various sorts is often [39,82], if not always [40], conscious. That

suggests a parallel question about mental states' being conscious independent of distinctively metacognitive processes. Is there any utility to a mental state's being conscious over and above that mental state's occurring without being conscious? And if so, what does that utility consist of?

Metacognition and consciousness are distinct phenomena; so metacognitive utility will not automatically carry over to the consciousness of mental states. But similarities between the two suggests that because metacognition is highly useful, consciousness may have some utility as well.

Metacognitive judgements are judgements about representational states, and mental states' being conscious arguably always involves one's being aware of those states. On a higher-order theory, mental states' being conscious consists of such awareness, and even on first-order theories a state's being conscious may typically facilitate such awareness. So it seems inviting to speculate that there will be some utility to mental states' being conscious beyond the utility of mental states that occur without being conscious.

And there are considerations apart from the analogy with metacognition that may lead us to expect that consciousness plays some useful role. We often seek to understand features of the psychology of people as well as other animals by appeal to some utility those features have. So it is natural to look for a positive role played by mental states' being conscious. The search for such utility may result from a focus on evolutionary selection pressures, but it need not; it is often productive, independent of issues about evolutionary origin, to look for some utility that aspects of psychological functioning exhibit.

But the theoretical hunch that we will find some utility must not mislead us. We often understand aspects of psychological functioning by appeal not to their utility, but simply to the factors that cause the psychological functioning under investigation. So we must hold open the possibility that mental states' being conscious adds little or no utility to what those states would have, even without being conscious. In that case, we would have to explain why many states are conscious by appeal to those factors that cause those states to be conscious.

There are two broad areas that theorists have examined as promising candidates for such utility, both suggested by the analogy with metacognition. One is internal to the individual, having to do with rationality generally, and more specifically with planning, inference and executive function. The other pertains to social interactions, including informing others what mental states one is in and other potential ways of facilitating social cohesion.

#### (a) *Utility owing to rationality*

Rationality is a matter of how one's thoughts, desires, goals and other cognitive as well as volitional states cohere with one another. Similarly, for the efficacious laying of plans and the drawing of inferences, and executive function; these are all a matter of how one's cognitive and volitional states interact. And the relevant interactions hinge on the intentional content

of those states. Rationality and inference depend entirely on rational connections among the contents of cognitive and volitional states, as does the efficacy of our plans. And executive function, which involves inhibitory control and the fine-tuning and adjusting of the volitional states that lead to behaviour, again hinges on the content of those volitional states.

The exclusive dependence of these processes on intentional content is reflected in the widespread recognition, despite other theoretical disagreements, that such content must involve causal connections among states with suitably related content. Otherwise, the interactions among such states on which rationality, planning and executive control must rely would remain mysterious. Content goes hand in hand with causal potential.

These processes do often occur consciously; we are frequently aware of cognitive and volitional states that constitute these processes. But cognitive and volitional states also often occur without being conscious. And because rationality and efficacy of planning and executive control all hinge on the intentional content of the relevant states, it is unclear why the states' being conscious would in any way enhance those processes.

Even when the processes do involve conscious states, the efficacy of the processes depends on the content of the relevant states, not on the independent property of the states' being conscious [83]. Because volitional states occur without being conscious before they come to be conscious [27,28], the same may happen with cognitive states. If so, consciousness will not have any significant utility in either case. This fits well with findings by Dijksterhuis *et al.* [84], controversial [85] but recently replicated by Usher *et al.* [86], on which decisions about complex matters often turn out better when they result from thought processes that are largely not conscious. Indeed, conscious monitoring of one's thought processes is often awkward and inefficient.

Adopting a traditional inner-sense picture of consciousness, on which awareness of our conscious states is perceptual, makes it tempting to hold that consciousness does play a beneficial role in these rationality and related processes. Perceiving things enables us to negotiate our way among them, pick them up and move them about in ways that suit our purposes. So a perceptual model of the way in which we are aware of our conscious states encourages the idea that being thus aware of those states should enable us to rearrange them in similarly advantageous ways.

A perceptual model of mental states' being conscious is unconvincing, independent of these considerations. Perceiving involves mental qualities that occur in response to perceptible properties and enable us to discriminate among them. But there are no higher-order mental qualities that figure in the way we are aware of our mental states; such awareness is not perceptual [11].

Independent of any such perceptual model, however, mental states are simply not the sorts of things we can rearrange to suit our purposes as we do with the physical objects we perceive. We do revise and adjust our thinking and our volitional states, and executive function facilitates fine-tuning of such states. But such revising and adjusting typically or always hinges

on interactions among the relevant states solely as a result of their content. We revise poor plans and irrational thinking because of the conflicts in content they lead to and the causal potential states have exclusively in virtue of their content. Awareness of the states is seldom, if ever, needed [83].

These considerations apply not only to cognitive and volitional states, but also to qualitative states, such as perceptual sensations. Mental qualities, on the account sketched in §5, have distinctive perceptual roles. And the utility a perceptual state has for the organism will be due mainly or even solely to that perceptual role. Because that perceptual role is independent of qualitative states' being conscious, we have no reason to think that their being conscious would add any utility beyond the utility they already have due to perceptual role.

As noted in §4, learning new words that draw more fine-grained distinctions among one's qualitative states sometimes results in one's being aware of those states in more fine-grained ways; one comes to be able to draw more distinctions among one's qualitative states solely in virtue of one's first-person access to those states. But there is again no reason to think that qualitative states' being conscious in respect of these more fine-grained differences would add significant utility. Differences among qualitative states already enable fine-grained perceptual discriminations independent of the states' being conscious in respect of those fine-grained differences.

So it is unclear what utility is added by the states' coming to be conscious in respect of those differences apart from a measure of aesthetic pleasure. And that pleasure may well occur without being conscious; people sometimes exhibit an aesthetic preference without conscious awareness of what it is in virtue of which they have that preference. So the added utility would consist, at most, of the pleasure's coming to be conscious, something arguably marginal in the context of one's overall psychological functioning.

Because many mental processes occur without conscious awareness, it is natural to look for some utility of conscious awareness by seeing which mental processes are invariably accompanied by such awareness [87]. But that is not sufficient. Greater signal strength may independently result in a state's being conscious and in some other effect, much as variations in atmospheric pressure cause changes both in weather and in barometer readings, effects unconnected except for having a common cause. Even when some useful psychological function always occurs consciously, we need evidence that conscious awareness is required for that function and is not simply an irrelevant by-product, an epiphenomenon, as that term is used in medical contexts.

Weiskrantz [25, ch. 7] notes that flexible thinking seems not to recruit the perceptual contents that fail to occur consciously in disorders such as blindsight, prosopagnosia and amnesia. And he suggests that a benefit of those perceptual contents' being conscious may be their availability for such flexible thinking. But such disorders involve deficits in mental processing in addition to the failure of some perceptual states to occur consciously; so the unavailability of

such states for flexible thinking may not be due to those states' not being conscious. In addition, individuals with these disorders are habituated to recruiting conscious perceptions in their flexible thinking, perhaps only because so much conscious perceiving is available. And with individuals for whom no relevant perceptions are conscious, those that are not conscious can figure in flexible behaviour [88].

It is important to note a potential ambiguity in the term 'function'. The question of concern here is whether mental states' being conscious has any utility for the organism beyond the occurrence of those states without their being conscious. But we also speak of function as simply the observable or discernible effect something has, as Cohen & Dennett [89] evidently do. Their argument that consciousness cannot be separated from function hinges on noting that conscious states can be reported [25,26,42] and that they typically have other behavioural effects. And as they argue, conscious states that cannot be detected also cannot be studied. But that by itself does not help show that mental states' being conscious adds any utility for the organism beyond the utility of those states' occurring without being conscious.

#### **(b) *Utility owing to social interaction***

The ability to report conscious states raises the second type of utility that theorists have explored for mental states' being conscious: the utility of enhancing social interactions. If I can tell you what states I am in, that will facilitate interactions between us. And if a mental state is not conscious, I will be unable to tell you about it [87, p. 539].

It is often useful to know what mental states others are in. But being told is not the only way to learn that. Suppose I think it is raining. One way you can learn that I think that is by my explicitly saying, 'I think it is raining'. And I cannot tell you that unless the thought I report is conscious; indeed, my saying, 'I think that it is raining' expresses my awareness of that thought, the HOA in virtue of which the thought is conscious.

But you could also learn that I think that it is raining not by my explicitly saying that I think it is, but by my simply saying instead, 'It is raining'. I need not tell you that I have that thought; I can just express my thought verbally. And if I do just say, 'It is raining', I do not thereby express my awareness of that thought; I convey the thought, but I do not explicitly mention it. So in this case my thought's being conscious plays no role in my letting you know what I think. Consciousness is not needed for the social utility of letting others know about our thoughts.

The difference between expressing and reporting seems small in the case of an assertoric thought, such as the thought that it is raining, small enough that it is sometimes overlooked [90]. But the difference is vivid in other cases. If one wonders whether it is raining, one expresses one's wondering simply by asking, 'Is it raining?', whereas one reports the state by saying assertorically, 'I wonder whether it is raining'. The difference stands out even more with emotions; typically, we learn of somebody's having some emotion from that person's expressions of

emotion, verbal and non-verbal, rather than from the person's explicitly saying, 'I am feeling angry [sad, happy, and so forth]' [1, ch. 11].

The most straightforward way to convey that one thinks something is to say whatever it is that one thinks. Similarly, with doubting, wondering and so forth, one can always simply express these states verbally.

But one can also move a step up, and explicitly say that one has the thought. We can describe this move from verbally expressing to explicitly reporting as psychological ascent, by analogy with the move from asserting a sentence to saying that the sentence is true, which Quine [91] calls 'semantic ascent'. And just as describing a sentence as true has no cognitive advantage over simply asserting the sentence, so explicitly saying that one has a particular thought has no utility beyond simply expressing the thought verbally.

Typically, when one verbally expresses a thought, that thought is conscious. But its being conscious is not what enables the thought to be expressed. Verbally expressed thoughts are typically conscious because our ability to describe our own thoughts induces a propensity to be aware of any thoughts one verbally expresses [1, ch. 10]. The thought's being conscious plays no role in enabling one to express the thought verbally.

Frith [87, p. 539] notes that there can be a distinctively social aspect to subjects' reporting conscious awareness of a stimulus. Reports are communicative acts, and expectations that experimenters or others have may sometimes influence those reports. Still, the reporting is not a joint endeavour; a report expresses an individual's awareness or a lack of awareness of a mental state, and so reflects whether particular states occur consciously.

Frith [87] has also developed a challenging view on which conscious awareness has a distinctively social utility that derives from the conscious awareness of one's own and others' agency. Conscious awareness of our own agency and that of others, he urges, enhances social cooperation. Distinguishing voluntary from non-voluntary behaviour on others is plainly important for cooperative social interactions, and we often distinguish those things consciously.

But young infants also draw that distinction [92,93], casting doubt on whether conscious awareness is required to do so; indeed, even newborn chicks seem to draw a distinction between animate and inanimate movement [94], suggesting that the discerning of voluntary behaviour may be too primitive to rely on conscious perception. So it is unclear what additional benefit there might be to doing so consciously. Perhaps Frith [87, p. 535] sees added utility because he holds that conscious awareness is required in deciding for oneself what to do. But it is also unclear that such decisions cannot occur without being conscious [63,64]; so conscious awareness may well add little, if any, utility even there.

It is sometimes held that our ability to tell what mental states others are in has some connection with the ability to be aware of our own mental states in the way we are when those states are conscious [95]. But the two abilities have little, if anything, in common beyond their both making use of relevant concepts of the relevant states. The mind-reading ability that

enables us to tell what states others are in relies on non-verbal behaviour, whereas the awareness of one's own conscious states is independent of self-observation. And although metacognitive judgements may enhance mind-reading abilities [96], both are independent of conscious awareness.

It is worth stressing that the considerations that undermine the widespread conviction that mental states' being conscious adds significant utility do not rest on one's having adopted a higher-order theory of what it is for mental states to be conscious. A number of the considerations that point to an absence of added utility also support higher-order theories. But those considerations lend support to higher-order theories, independent of casting doubt on there being significant added utility to a state's being conscious.

It may strike many as surprising, for reasons noted at the outset of the section, that there should be no significant utility due exclusively to mental states' being conscious, utility that does not occur when those states occur without being conscious and is not an independent by-product of the processes that lead to their being conscious. But theory and empirical investigation seem so far not to have uncovered any such benefit.

## REFERENCES

- Rosenthal, D. M. 2005 *Consciousness and mind*. Oxford, UK: Clarendon Press.
- Lau, H. 2008 Are we studying consciousness yet? In *Frontiers of consciousness: Chichele lectures* (eds L. Weiskrantz & M. Davies), pp. 245–258. Oxford, UK: Oxford University Press.
- Locke, J. 1975/1700 *An essay concerning human understanding* (ed. P. H. Nidditch). Oxford, UK: Oxford University Press.
- Kant, I. 1998/1787 *Critique of pure reason* (eds P. Guyer & A. W. Wood). Cambridge, UK: Cambridge University Press.
- Armstrong, D. M. 1980 What is consciousness? In *The nature of mind* (D. M. Armstrong), pp. 55–67. St. Lucia, Queensland: University of Queensland Press.
- Lycan, W. G. 1996 *Consciousness and experience*. Boston, MA: MIT Press.
- Weisberg, J. 2010 Misrepresenting consciousness. *Philos. Stud.* **154**, 409–433. (doi:10.1007/s11098-010-9567-3)
- Weisberg, J. 2011 Abusing the notion of what-it's-like-ness: a response to Block. *Analysis* **71**, 438–443. (doi:10.1093/analys/anr040)
- Brentano, F. 1973/1874 *Psychology from an empirical standpoint* (eds A. C. Rancurello, D. B. Terrell & L. L. McAlister), pp. 126–133. New York, NY: Routledge & Kegan Paul.
- Kriegel, U. 2009 *Subjective consciousness: a self-representational theory*. Oxford, UK: Oxford University Press.
- Rosenthal, D. M. 2004 Varieties of higher-order theory. In *Higher-order theories of consciousness* (ed. R. J. Gennaro), pp. 19–44. Amsterdam, The Netherlands: John Benjamins.
- Dretske, F. I. 1993 Conscious experience. *Mind* **102**, 263–283. (doi:10.1093/mind/102.406.263)
- Dretske, F. I. 2006 Perception without awareness. In *Perceptual experience* (eds T. S. Gendler & J. Hawthorne), pp. 147–180. Oxford, UK: Oxford University Press.
- Block, N. 1995 On a confusion about a function of consciousness. *Behav. Brain Sci.* **18**, 227–247. (doi:10.1017/S0140525X00038188)
- Baars, B. 1998 *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.

- 16 Dehaene, S. & Naccache, L. 2001 Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37. (doi:10.1016/S0010-0277(00)00123-2)
- 17 Dehaene, S., Sergent, C. & Changeux, J. P. 2003 A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl Acad. Sci. USA* **100**, 8520–8525. (doi:10.1073/pnas.1332574100)
- 18 Block, N. 2009 Comparing the major theories of consciousness. In *The cognitive neurosciences IV* (ed. M. Gazzaniga), pp. 1111–1122. Cambridge, MA: MIT Press.
- 19 Block, N. 2007 Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* **30**, 481–548. (doi:10.1017/S0140525X07002786)
- 20 Block, N. 1978 Troubles with functionalism. In *Minnesota studies in the philosophy of science*, vol. IX (ed. C. W. Savage), pp. 261–325. Minneapolis, MN: University of Minnesota Press.
- 21 Lau, H. & Passingham, R. E. 2006 Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl Acad. Sci. USA* **103**, 18 763–18 768. (doi:10.1073/pnas.0607716103)
- 22 Lau, H. & Rosenthal, D. 2011 Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* **15**, 365–373. (doi:10.1016/j.tics.2011.05.009)
- 23 Block, N. 2011 The Anna Karenina theory of the unconscious. *Neuropsychanalysis* **13**, 34–37.
- 24 Block, N. 2011 The higher order approach to consciousness is defunct. *Analysis* **71**, 419–431. (doi:10.1093/analys/anr037)
- 25 Weiskrantz, L. 1997 *Consciousness lost and found: a neuro-psychological exploration*. Oxford, UK: Oxford University Press.
- 26 Weiskrantz, L. 1998 Consciousness and commentaries. *Int. J. Psychol.* **33**, 227–233. (doi:10.1080/002075998400411)
- 27 Libet, B., Gleason, C. A., Wright Jr, E. W. & Pearl, D. K. 1983 Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). *Brain* **106**, 623–642. (doi:10.1093/brain/106.3.623)
- 28 Haggard, P. 1999 Perceived timing of self-initiated actions. In *Cognitive contributions to the perception of spatial and temporal events* (eds G. Aschersleben, T. Bachmann & J. Müsseler), pp. 215–231. Amsterdam, The Netherlands: Elsevier.
- 29 Rosenthal, D. 2002 The timing of conscious states. *Conscious. Cogn.* **11**, 215–220. (doi:10.1006/ccog.2002.0558)
- 30 Koriat, A. 2007 Metacognition and consciousness. In *The Cambridge handbook of consciousness* (eds P. D. Zelazo, M. Moscovitch & E. Thompson), pp. 289–325. Cambridge, UK: Cambridge University Press.
- 31 Nelson, T. O. & Narens, L. 1994 Why investigate metacognition? In *Metacognition: knowing about knowing* (eds J. Metcalfe & A. P. Shimamura), pp. 1–25. Cambridge, MA: MIT Press.
- 32 Nelson, T. (ed.) 1992 *Metacognition: core readings*. Needham Heights, MA: Allyn and Bacon.
- 33 Metcalfe, J. & Shimamura, A. P. (eds) 1994 *Metacognition: knowing about knowing*. Cambridge, MA: MIT Press.
- 34 Rosenthal, D. 2000 Consciousness, content, and metacognitive judgments. *Conscious. Cogn.* **9**, 203–214. (doi:10.1006/ccog.2000.0437)
- 35 Levin, D. T., Momen, N., Drivdahl, S. B. & Simons, D. J. 2000 Change blindness blindness: the metacognitive error of overestimating change-detection ability. *Vis. Cogn.* **7**, 397–412. (doi:10.1080/135062800394865)
- 36 Overgaard, M. & Sandberg, K. 2012 Kinds of access: different methods for report reveal different kinds of metacognitive access. *Phil. Trans. R. Soc. B* **367**, 1287–1296. (doi:10.1098/rstb.2011.0425)
- 37 Descartes, R. 1984/1641 *The philosophical writings of Descartes*, vol. 2 (eds J. Cottingham, R. Stoothoff & D. Murdoch). Cambridge, UK: Cambridge University Press.
- 38 Block, N. 2011 Response to Rosenthal and Weisberg. *Analysis* **71**, 443–448. (doi:10.1093/analys/anr036)
- 39 Fernandez-Duque, D., Baird, J. A. & Posner, M. I. 2000 Awareness and metacognition. *Conscious. Cogn.* **9**, 324–326. (doi:10.1006/ccog.2000.0449)
- 40 Kentridge, R. W. & Heywood, C. A. 2000 Metacognition and awareness. *Conscious. Cogn.* **9**, 308–312. (doi:10.1006/ccog.2000.0448)
- 41 Seth, A. K. 2008 Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious. Cogn.* **17**, 981–983. (doi:10.1016/j.concog.2007.05.008)
- 42 Marcel, A. J. 1983 Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cognit. Psychol.* **15**, 238–300. (doi:10.1016/0010-0285(83)90010-5)
- 43 Dienes, Z. 2008 Subjective measures of unconscious knowledge. *Prog. Brain Res.* **168**, 49–64. (doi:10.1016/S0079-6123(07)68005-4)
- 44 Persaud, N., McLeod, P. & Cowey, A. 2007 Post-decision wagering objectively measures awareness. *Nat. Neurosci.* **10**, 257–261. (doi:10.1038/nn1840)
- 45 Koch, C. & Preusschoff, K. 2007 Betting the house on consciousness. *Nat. Neurosci.* **10**, 140–141. (doi:10.1038/nn0207-140)
- 46 Schurger, A. & Sher, S. 2008 Awareness, loss aversion, and post-decision wagering. *Trends Cogn. Sci.* **12**, 209–210. (doi:10.1016/j.tics.2008.02.012)
- 47 Dienes, Z. & Seth, A. 2010 Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious. Cogn.* **19**, 674–681. (doi:10.1016/j.concog.2009.09.009)
- 48 Mellor, D. H. 1971 *The matter of chance*. Cambridge, UK: Cambridge University Press.
- 49 van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S. & Lamme, V. A. F. 0000 Frontal cortex mediates unconsciously triggered inhibitory control. *J. Neurosci.* **28**, 8053–8062. (doi:10.1523/JNEUROSCI.1278-08.2008)
- 50 Schnitzspahn, K. M., Zeintl, M., Jäger, T. & Kliegel, M. 2011 Metacognition in prospective memory: are performance predictions accurate? *Can. J. Exp. Psychol.* **65**, 19–26. (doi:10.1037/a0022842)
- 51 Paynter, C. A., Reder, L. M. & Kieffaber, P. D. 2009 Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia* **47**, 796–803. (doi:10.1016/j.neuropsychologia.2008.12.009)
- 52 Pieschl, S. 2009 Metacognitive calibration: an extended conceptualization and potential applications. *Metacognit. Learn.* **4**, 3–31. (doi:10.1007/s11409-008-9030-4)
- 53 Scholl, B. J., Simons, D. J. & Levin, D. T. 2004 Change blindness blindness: an implicit measure of a metacognitive error. In *Thinking and seeing: visual metacognition in adults and children* (ed. D. T. Levin), pp. 145–163. Cambridge, MA: MIT Press.
- 54 Nagel, T. 1974 What is it like to be a bat? *Philos. Rev.* **83**, 435–450. (doi:10.2307/2183914)
- 55 Levine, J. 2001 *Purple haze: the puzzle of consciousness*. New York, NY: Oxford University Press.
- 56 Neander, K. 1998 The division of phenomenal labor: a problem for representational theories of consciousness. *Phil. Perspect.* **12**, 411–434. (doi:10.1111/0029-4624.32.512.18)

- 57 Wilberg, J. 2010 Consciousness and false HOTs. *Philos. Psychol.* **23**, 617–638. (doi:10.1080/09515089.2010.514567)
- 58 Dennett, D. C. 1991 *Consciousness explained*. Boston, MA: Little, Brown.
- 59 Ro, T., Singhal, N. S., Breitmeyer, B. G. & Garcia, J. O. 2009 Unconscious processing of color and form in meta-contrast masking. *Atten. Percept. Psychol.* **71**, 95–103. (doi:10.3758/APP.71.1.95)
- 60 Breitmeyer, B. G., Ro, T. & Singhal, N. S. 2004 Unconscious color priming occurs at stimulus—not percept—dependent levels of processing. *Psychol. Sci.* **15**, 198–202. (doi:10.1111/j.0956-7976.2004.01503009.x)
- 61 Breitmeyer, B. G., Ro, T., Ögmen, H. & Todd, S. 2007 Unconscious, stimulus-dependent priming and conscious, percept-dependent priming with chromatic stimuli. *Percept. Psychophys.* **69**, 550–557. (doi:10.3758/BF03193912)
- 62 Grimes, J. 1996 On the failure to detect changes in scenes across saccades. In *Perception* (ed. K. Akins), pp. 89–110. New York, NY: Oxford University Press.
- 63 Desmurget, M. & Sirigu, A. 2009 A parietal-premotor network for movement intention and motor awareness. *Trends Cogn. Sci.* **13**, 411–419. (doi:10.1016/j.tics.2009.08.001)
- 64 Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C. & Sirigu, A. 2009 Movement intention after parietal cortex stimulation in humans. *Science* **324**, 811–813. (doi:10.1126/science.1169896)
- 65 Wegner, D. M. 2002 *The illusion of conscious will*. Cambridge, MA: MIT Press.
- 66 Pérez-Carpinell, J., Baldoví, R., de Fez, M. D. & Castro, J. 1998 Color memory matching: time effect and other factors. *Color Res. Appl.* **23**, 234–247. (doi:10.1002/(SICI)1520-6378(199808)23:4<234::AID-COL8>3.0.CO;2-P)
- 67 Halsey, R. M. & Chapanis, A. 1951 On the number of absolutely identifiable spectral hues. *J. Opt. Soc. Am.* **41**, 1057–1058. (doi:10.1364/JOSA.41.001057)
- 68 Seashore, C. E. 1967 *The psychology of music*. New York, NY: Dover Publications.
- 69 van Boxtel, J. J. A., Tsuchiya, N. & Koch, C. 2010 Opposing effects of attention and consciousness on after-images. *Proc. Natl Acad. Sci. USA* **107**, 8883–8888. (doi:10.1073/pnas.0913292107)
- 70 van Gaal, S. & Fahrenfort, J. J. 2008 The relationship between visual awareness, attention, and report. *J. Neurosci.* **28**, 5401–5402. (doi:10.1523/JNEUROSCI.1208-08.2008)
- 71 Schurger, A., Cowey, A., Cohen, J., Treisman, A. & Tallon-Baudry, C. 2008 Distinct and independent correlates of awareness and attention in a hemianopic patient. *Neuropsychologia* **46**, 2189–2197. (doi:10.1016/j.neuropsychologia.2008.02.020)
- 72 Koch, C. & Tsuchiya, N. 2007 Attention and consciousness: two distinct brain processes. *Trends Cogn. Sci.* **11**, 16–22. (doi:10.1016/j.tics.2006.10.012)
- 73 Kentridge, R. W., Heywood, C. A. & Weiskrantz, L. 2004 Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* **42**, 831–835. (doi:10.1016/j.neuropsychologia.2003.11.001)
- 74 Rosenthal, D. 2011 Exaggerated reports: reply to Block. *Analysis* **71**, 431–437. (doi:10.1093/analys/anr039)
- 75 Nisbett, R. E. & Wilson, T. D. 1977 Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259. (doi:10.1037/0033-295X.84.3.231)
- 76 Johansson, P., Hall, L., Sikström, S. & Olsson, A. 2005 Failure to detect mismatches between intention and outcome in a simple decision task. *Science* **310**, 116–119. (doi:10.1126/science.1111709)
- 77 Frith, C. 2007 *Making up the mind: how the brain creates our mental world*. Malden, MA: Blackwell Publishing.
- 78 Koyama, T., McHaffie, J. G., Laurienti, P. J. & Coghill, R. C. 2005 The subjective experience of pain: where expectations become reality. *Proc. Natl Acad. Sci. USA* **102**, 12 950–12 955. (doi:10.1073/pnas.0408576102)
- 79 Strawson, G. 2006 Realistic monism: why physicalism entails panpsychism. *J. Conscious. Stud.* **13**, 3–31.
- 80 Rosenthal, D. 2010 How to think about mental qualities. *Philos. Issues* **20**, 368–393. (doi:10.1111/j.1533-6077.2010.00190.x)
- 81 Billock, V. A., Gleason, G. A. & Tsou, B. H. 2001 Perception of forbidden colors in retinally stabilized equiluminant images: an indication of softwired cortical color opponency? *J. Opt. Soc. Am. A* **18**, 2398–2403. (doi:10.1364/JOSAA.18.002398)
- 82 Fernandez-Duque, D., Baird, J. A. & Posner, M. I. 2000 Executive attention and metacognitive regulation. *Conscious. Cogn.* **9**, 288–307. (doi:10.1006/ccog.2000.0447)
- 83 Rosenthal, D. M. 2008 Consciousness and its function. *Neuropsychologia* **46**, 829–840. (doi:10.1016/j.neuropsychologia.2007.11.012)
- 84 Dijksterhuis, A., Bos, M. W., Nordgren, L. F. & van Baaren, R. B. 2006 On making the right choice: the deliberation-without-attention effect. *Science* **311**, 1005–1007. (doi:10.1126/science.1121629)
- 85 Waroquier, L., Marchiori, D., Klein, O. & Cleeremans, A. 2010 Is It better to think unconsciously or to trust your first impression? A reassessment of unconscious thought theory. *Soc. Psychol. Perspect. Sci.* **1**, 111–118. (doi:10.1177/1948550609356597)
- 86 Usher, M., Russo, Z., Weyers, M., Brauner, R. & Zakay, D. 2011 The impact of the mode of thought in complex decisions: intuitive decisions are better. *Front. Psychol.* **2**, 37. (doi:10.3389/fpsyg.2011.00037)
- 87 Frith, C. 2010 What is consciousness for? *Pragmat. Cognit.* **18**, 497–551. (doi:10.1075/pc.18.3.03fri)
- 88 de Gelder, B., Tamietto, M., van Boxtel, G., Rainer Goebel, R., Sahraie, A., van den Stock, J., Stienen, B. M. C., Weiskrantz, L. & Pegna, A. 2008 Intact navigation skills after bilateral loss of striate cortex. *Curr. Biol.* **18**, R1128–R1129. (doi:10.1016/j.cub.2008.11.002)
- 89 Cohen, M. A. & Dennett, D. C. 2011 Consciousness cannot be separated from function. *Trends Cogn. Sci.* **15**, 358–364. (doi:10.1016/j.tics.2011.06.008)
- 90 Wittgenstein, L. 2009 *Philosophical Investigations*, 4th edn. (eds P. M. S. Hacker & J. Schulte), II, x, 190–192. Oxford, UK: Wiley-Blackwell.
- 91 Quine, W. V. 1960 *Word and object*. Cambridge, MA: MIT Press.
- 92 Tomasello, M. 2000 Culture and cognitive development. *Curr. Dir. Psychol. Sci.* **9**, 37–40. (doi:10.1111/1467-8721.00056)
- 93 Behne, T., Carpenter, M., Call, J. & Tomasello, M. 2005 Unwilling versus unable: infants' understanding of intentional action. *Dev. Psychol.* **41**, 328–337. (doi:10.1037/0012-1649.41.2.328)
- 94 Mascalzonia, E., Regolina, L. & Vallortigara, G. 2010 Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proc. Natl Acad. Sci. USA* **107**, 4483–4485. (doi:10.1073/pnas.0908792107)
- 95 Carruthers, P. 2000 *Phenomenal consciousness: a naturalistic theory*. Cambridge, UK: Cambridge University Press.
- 96 Koriat, A. & Ackerman, R. 2010 Metacognition and mindreading: judgments of learning for self and other during self-paced study. *Conscious. Cogn.* **19**, 251–264. (doi:10.1016/j.concog.2009.12.010)