In: Gazzaniga MS (ed.) *The Cognitive Neurosciences*, pp. 1295–1306. Cambridge, MA: MIT Press.
Lumer ED (2000) Binocular rivalry and human visual awareness. In: Metzinger T (ed.) *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, pp. 231–240. Cambridge, MA: MIT Press.
Marcel AJ (1983) Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology* **15**: 197–237.
Norman DA and Shallice T (1986) Attention to action: willed and automatic control of behavior. In: Davidson RJ, Schwartz GE and Shapiro D (eds) *Consciousness and Self-regulation*, vol. 4, pp. 1–18. New York, NY: Plenum Press.
Posner MI (1994) Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences of the USA* **91**: 7398–7403.
Posner MI and Rothbart MK (1991) Attentional mechanisms and conscious experience. In: Milner AD and Rugg MD (eds) *The Neuropsychology of Consciousness*, pp. 91–111. London, UK: Academic Press.

Consciousness and Higher-order Thought

Intermediate article

David M Rosenthal, City University of New York Graduate School, New York, New York, USA

CONTENTS

The higher-order-thought hypothesis is a proposed explanation of what it is for a mental state to be a conscious state and hence of how conscious mental states differ from mental states that are not conscious.

INTRODUCTORY

Any satisfactory theoretical treatment of consciousness must begin by distinguishing several phenomena to which the term 'consciousness' applies. We describe people and other animals as being conscious when they are awake and responsive to sensory stimulation. What it is for a creature to be conscious in this sense is primarily a biological matter and peripheral to cognitive science and related theory.

We also describe creatures as being *conscious of* various things, for example, when they sense those things or think about them as being present. Sensing and thinking are central to cognitive functioning, but their nature is not what theorists typically have in mind in discussing consciousness. Rather, theorists have in mind primarily a third application of the term 'consciousness', by which we describe thinking and sensing itself as being conscious or not. It is this third use which dominates theoretical discussion about consciousness. The central issue is what it is for a mental state, such as thinking, sensing, and feeling, to be conscious, and more specifically what distinguishes the conscious cases from those which are not.

THEORIES OF CONSCIOUSNESS

It is fundamental to a mental state's being conscious that the individual in the state is aware of being in it. This is clear from consideration of mental states an individual is unaware of. If somebody is altogether unaware of thinking, feeling, or sensing something, that thinking, feeling or sensing does not count as conscious. Part of what it is for a state to be conscious is that one is conscious of being in that state.

Some theorists deny this, arguing that we are never conscious of our conscious states (Searle,

1992), or at least that conscious states occur without one's being conscious of them (Dretske, 1995). Thus Dretske, for example, urges that a state's being conscious consists not in one's being conscious of it, but in one's being conscious of something by virtue of being in that state.

This view has a disadvantage. Since sensing and thinking about things typically make one conscious of them, such states could not, on this view, occur without being conscious. Such theorists accordingly argue that the usual examples given of mental states that are not conscious are unconvincing. One especially common type of example does seem vulnerable to this charge. Armstrong (1978/1980) and others have appealed to the case of the long-distance driver who seems for a time not to notice the road consciously. But it may be that the driver notices the road consciously but simply does not at all remember doing so.

There are, however, other examples of mental states that more indisputably occur without being conscious. People often act in ways that betray some feeling, or belief, or desire of which they are wholly unaware until it is pointed out to them; this even happens with pains that are revealed by gestures or bodily movements. And people sometimes respond in a very fine-grained way to things that occur so far in the periphery of their visual field that they have no conscious perception of them.

Many experimental results confirm these commonsense observations (Merikle *et al*., 2001). In masked-priming experiments, subjects presented very briefly with two successive stimuli report not being aware of the first at all, even though that stimulus has a demonstrable effect on mental processing (e.g. Marcel, 1983a, 1983b). And blindsight subjects, in whom part of the primary visual cortex has been destroyed, deny seeing visual stimuli in the relevant area of the visual field, though they can be prompted to guess the visible characteristics of such stimuli with startlingly high accuracy (Weiskrantz, 1997). Though conscious sensing is absent in these cases, subjects' behavior indicate the occurrence of sensing that is not conscious.

Some theorists have argued that it is circular to explain a mental state's being conscious in terms of an individual's being conscious of that state, since that would be to explain consciousness by appeal to consciousness (e.g. Goldman, 1993). But that explanation is not circular. Being conscious of something is sensing it or thinking about it as present. And, since we understand what it is to sense something or think about it even when that sensing or thinking is not conscious, we understand what it is to be conscious *of* something independently of knowing what it is for mental states to be conscious.

Even if a state's being conscious consists in one's being conscious of that state, theories divide about just how one is conscious of one's conscious states. The traditional and most widespread view is that one senses or perceives one's conscious states. But thinking about something can also make one conscious of that thing, and an alternative theory has been developed on which we do not sense our conscious states, but instead are conscious of them by having thoughts about them. It is useful to refer to the thoughts or sensations in virtue of which one is conscious of one's mental states as 'higher-order thoughts' or 'higher-order sensations'.

THE INNER-SENSE MODEL

The idea that we sense our conscious states has a long history. Locke (1700/1975) speaks of an 'internal sense' by which we are conscious of our mental states, and Kant (1787/1998) speaks of 'inner sense.' More recently, the idea has been defended by Armstrong (1978/1980) and by Lycan (1996).

Several factors suggest an account in terms of such higher-order sensing. For one thing, nothing seems to mediate between the the things we sense and our sensing them. And this intuitively unmediated character of sensing might explain why the way we are aware of our conscious states seems to be direct and immediate.

Another factor has to do with the qualitative character of conscious sensory experience. That qualitative character enters our mental lives through sensing; thinking has no qualitative character. So it may seem that the only way we could be conscious of this qualitative aspect of experience is by sensing it. A third source of the idea that we sense our conscious states is the sense we have that we are regularly and reliably conscious of many of our own mental states. And the best explanation for this may be that we monitor our mental states in the way that our exteroceptive senses monitor the environment (Armstrong, 1978/1980; Lycan, 1996).

But these considerations are far from decisive. Although nothing seems to mediate between our mental states and our consciousness of them, we need not appeal to higher-order sensing to explain that appearance of immediacy. Having thoughts about our mental states would also make us conscious of those states, and if we aware of nothing mediating between those thoughts and the states they are about, our consciousness of those states would also seem to be unmediated.

Perhaps some monitoring mechanism in the brain does subserve our being conscious of many of our mental states, but monitoring need not be sensory. The brain monitors many bodily functions in ways that do not at all resemble sensing. What differentiates sensing from other processes are the distinctive qualitative properties that occur when we sense. When sensing is conscious, we are conscious of these distinguishing qualities, qualities that vary with what is sensed, though these qualities also occur without our being at all aware of them.

The third consideration that seemed to support the inner-sense model, namely, the qualitative character of sensing, actually provides a compelling reason to reject the model (Rosenthal, 1997). Although sensations and perceptual states exhibit distinguishing qualitative properties, the way we are conscious of our own mental states does not. This is evident when the states we are conscious of are thoughts, beliefs, desires, and other so-called intentional states; these states have no qualitative properties, and there is no qualitative character to the way we are conscious of them. But even when the states we are conscious of are qualitative, as with our sensations and emotions, the qualities belong to the states we are conscious *of*, not to the way we are conscious of them.

Some theorists describe the inner-sense model in terms of higher-order perceiving of mental states (Güzeldere, 1995). Since perceiving, like sensing, has qualitative character, a higher-order perception view faces the difficulty that no higher-order qualities occur. But perceiving not only has qualitative character, but also resembles thinking in having conceptual content. So, if we had higher-order perceptions of our mental states, we would still need to determine whether the qualities or conceptual content of the perceptions were responsible for our being conscious of our mental states. Compare Güzeldere (1995), who argues that the higher-order-perception model collapses into a model that invokes higher-order thoughts.

THE HIGHER-ORDER-THOUGHT MODEL

The two ways of being conscious of things are sensing them and having thoughts about them as being present. Since we are not conscious of our mental states by sensing them, the best explanation of how we are conscious of some of our mental states is that we have higher-order thoughts (HOTs) about them (Rosenthal, 1986, 1993; in press a, b). It would be explanatorily empty to insist that we are conscious of them in some third way unless we have an independent grasp of what that third way consists in.

Difficulties with the inner-sense view actually suggest the HOT model. The higher-order states in virtue of which we are conscious of our mental states lack qualitative properties, and a thought that something is present makes one conscious of that thing in a way that involves no higher-order qualities. If there is a brain mechanism that monitors mental states, it might well make one conscious of those states by producing HOTs about them. And, if those HOTs seemed to arise independently of any inference, it would seem subjectively as though one is conscious of one's mental states in a way that is direct and unmediated.

It is important to distinguish between the ordinary way in which mental states are conscious and the focused, reflective way in which states can become conscious when we introspect them. The HOT model affords a natural explanation of this difference. The HOTs in virtue of which we are conscious of our mental states in ordinary, nonintrospective cases are not, themselves, conscious thoughts; HOTs make one aware of various mental states, but without one's being conscious also of the HOTs themselves. When one introspects a state, one deliberately focuses attention on it. One thereby becomes aware not only of the introspected state, but also of one's being conscious of it. So in these cases the relevant HOTs are themselves conscious thoughts (Rosenthal, 2000a).

Because HOTs need not be conscious, and indeed usually are not, people will normally be unaware of their presence. The occurrence of HOTs is not established by our being aware of them, since we are conscious of them only in the special case of introspection. HOTs are theoretical posits whose occurrence is established by theoretical considerations of the sort sketched above.

These considerations help dispel a certain misunderstanding. It is sometimes held (e.g. Block, 1995a; Chalmers, 1996) that the HOT model explains only introspective consciousness. That would be so if the model appealed only to conscious HOTs, establishing their occurrence by way of subjects' reports. But the HOTs the model invokes typically are not conscious, and they are intended to explain ordinary, nonintrospective consciousness.

A HOT is a thought to the effect that one is in a particular state, and so makes reference to oneself. Such reference to the self does not require any sophisticated concept of the self, but only a concept strong enough to distinguish oneself from

everything else and to form thoughts about the self thus distinguished (Rosenthal, in press c). Nor do HOTs need to describe their target states in terms of some concept of the mind; they can describe those states simply in terms of their role in perceiving, or thinking or in information-processing terms.

VARIANT HIGHER-ORDER-THOUGHT THEORIES

A number of variants HOT theories have been put forth. Some differ in only minimal ways from the hypothesis just described. For example, Mellor (1977–78) appeals to second-order beliefs to explain only what it is for beliefs to be conscious, and denies that such an explanation works for any other types of mental state. But the foregoing arguments apply equally to all types of mental state. And Rolls (1998) has argued for a model on which the HOTs must be linguistic in character. Since Rolls construes being linguistic to cover any syntactically composite mode of representation, this again is at most a slight modification of the model.

Brentano (1874/1973) argued that the higher-order state in virtue of which one is conscious of a mental state is internal to the target state in question. Brentano's examples are largely perceptual, which leads Brentano to see the higher-order state as being perceptual as well; so his view may be best construed as a variant of the higher-order sensing model. Others have advanced views, however, that posit HOTs that are internal to their targets (Kobes, 1996; Gennaro, 1996).

But no view on which HOTs are internal to their targets is likely to succeed. Intentional states are individuated not only by their content but also by mental attitudes, such as mental affirmation, doubt, wonder, hope, and the like. Just as no single state can have two distinct contents, so no single state can exhibit two distinct mental attitudes. The mental attitude of HOTs is always assertoric; HOTs affirm that one is in a particular state. But, if HOTs were internal to their targets, then a conscious case of wondering something would exhibit the mental attitudes both of wondering and of affirming. So HOTs cannot in general be part of their targets. Similar considerations apply to Brentano's perceptual variant, since every perceptual state belongs to some sensory modality, and none of those standard modalities is suitable for making one conscious of one's mental states.

On another variant of the model developed by Carruthers (1996, 2000), a state need not be the object of an actual HOT to be conscious; it is enough that the state simply be *disposed* to cause a HOT about it. One main motive for this variant is that it avoids the high cost, both in computational capacity and cognitive space, of having actual HOTs for each of one's conscious states.

But that consideration is not all that compelling. HOTs very likely take less to implement cortically than their targets, since a HOT simply represents one as being in a particular state; so their causal connections will likely be far less complex than those of the perceptual and cognitive states the HOTs are about. And, since cortical capacity is known to be far from fully utilized, the cost of implementing actual HOTs is unlikely to be significant. Introspection makes this objection seem more pressing than it is. Because we are never conscious of many thoughts at once, it seems that we could not have many HOTs. But HOTs are seldom conscious, and we could have many at once that are not conscious. And introspection cannot be a reliable guide to the mind's nonconscious operations.

The principal reason for a higher-order account is to explain how we are conscious of all our conscious states. But being disposed to have a thought about something does not make one conscious of that thing. So the question arises about how a state's being disposed to cause a HOT could result in one's being conscious of that state.

Carruthers's answer appeals to a particular theory of mental content. On that theory, the content a state has is a matter of what it is disposed to cause. So Carruthers argues that a state's simply being disposed to cause a HOT can confer suitable higher-order content on the state itself. Both teleological (e.g. Millikan, 1984) and inferential-role (e.g. Block, 1986; Peacocke, 1992) theories of content might allow for this result. A state's having such higher-order content directed upon itself would then explain why one is conscious of being in that state.

This reply faces several difficulties. For one thing, these theories of content are far from uncontroversial, and it is preferable to have one's theory of consciousness committed to as little as possible that is not widely accepted. Moreover, since the higher-order content in virtue of which one is conscious of the state is internal to the state itself, the dispositional theory would face the difficulty about mental attitudes that faces any model on which the higher-order state is internal to its target.

Most important, any state with suitable first-order content would, on this model, have dispositional properties that result in its having higher-order content. So one would be conscious of any state that had that first-order content. Since a

state's being conscious or not would depend on its first-order content, the model seems unable to explain how states of a given type can sometimes be conscious and sometimes not.

HIGHER-ORDER THOUGHTS AND SPEECH

It is widely accepted that, given a creature with suitable linguistic capacities, a mental state's being conscious coincides with that creature's ability to report noninferentially that it is in that state. Indeed, it is likely that this ability to report mental states noninferentially is what underlies the traditional intuition that we have special access to our mental states (Sellars, 1963).

This fits well with the HOT hypothesis. A noninferential report that one is in some mental state expresses one's thought that one is in that state, a thought that seems subjectively to rely on no inference. And it is arguable that the best explanation of this ability to report one's mental states noninferentially is that one actually has the HOTs that those reports would express (Rosenthal, 1993).

Some theorists hold that we cannot introspectively seem to be in a state that we are not in (e.g. Nelkin, 1996). Moreover, the seemingly noninferential character of our HOTs may suggest that they reflect some special access we have to our mental states. But thoughts need not be accurate, and thoughts about one's own mental states are no exception. Our consciousness even of what states we are in can be erroneous.

This is evident from compelling experimental findings in which subjects report thoughts and desires that they do not actually have. As with reports of thoughts and desires that do occur, these confabulations tend to make *ex post facto* sense of subjects' behavior, by rationalizing that behavior or by conforming to expectations or preconceived ideas. But in these cases evidence exists that subjects do not actually have the thoughts and desires they report (Nisbett and Wilson, 1977). Such confabulation appears to happen even with qualitative states, such as bodily or perceptual sensations (Staats *et al*., 1998; Holmes and Frost, 1976).

These findings again fit well with the HOT model. When one confabulates being in some mental state, one is conscious *of oneself as being in that state*. And consciousness is a matter of how one appears to oneself. So, if one has a HOT that represents one as being in some state, there is nothing subjectively, from the point of view of consciousness, that could enable one to tell whether any such state actually occurs.

When one thinks that something has a certain property, one in effect interprets that thing as having the property. So having a noninferential HOT that one is in some mental state amounts to spontaneously interpreting oneself as being in that state. This echoes Dennett's (1991) interpretivist account of consciousness. But Dennett (1987, 1991) holds that one's being in a mental state at all, independent of whether that state is conscious, is a matter of one's being subject to some appropriate interpretation. The HOT model does not endorse that more general view.

Because conscious states are sometimes confabulated, the states one is conscious of oneself as being in do not always exist. So we cannot describe a conscious state as a state that bears some actual relation to a HOT. Rather, a state's being conscious must consist in its being the *intentional object* of a HOT, the object that the thought seems to be about. And, because the state may not actually occur, we also cannot require that it cause the HOT.

OBJECTIONS

The HOT hypothesis is sometimes seen as a claim about the *concept* of a mental state's being conscious (Goldman, 2000). Construed as a hypothesis about conceptual analysis the hypothesis is implausible, since it seems *conceivable* that a state accompanied by a HOT could fail to be conscious (Balog, 2000; Rey, 2000). But the HOT model is best taken not as a conceptual claim, but as an empirical hypothesis about the nature of consciousness. On that construal, though we can conceive of a state's being accompanied by a HOT without being conscious, it turns out empirically that this never happens. One might also object that any specification of the nature of consciousness purports to state a metaphysical necessity, and the HOT hypothesis is not metaphysically necessary. But, even apart from the difficulty of determining what is metaphysically necessary in a way that is not question begging, it is arguable that the HOT hypothesis is necessary if at all only in the way in which truths of natural science are.

It is sometimes argued that the stipulation that HOTs be noninferential is arbitrary, since it should not matter to a state's being conscious whether the accompanying HOT is caused by an inference (Byrne, 1997; Seager, 1999). But the aetiology of the HOT does not matter, only the appearance of aetiology. A state is conscious only if we are conscious of it in a way that *seems* spontaneous and noninferential. As long as it seems that way, it does not matter how it is caused. Nor is there any

problem about establishing a causal or other connection between HOTs and their targets, as Natsoulas (1993) argues, since the targets are simply whatever states the HOTs are about.

Having a thought about something normally has no effect on it, and in particular does not make that thing conscious. So it may be objected that having a thought about a mental state could not result in that state's changing from not being conscious to being conscious (Block, 1995b; Byrne, 1997; Rey, 2000). But when a state becomes conscious that is not a change in the state itself, but only in whether one is noninferentially conscious of it; being conscious is not an intrinsic property of mental states.

Still, an objector might persist, since having a noninferential thought about a physical object does not result in that object or state's being conscious, why should having a HOT about a mental state result in that state's being conscious? But the only way objects might be conscious is the way a creature can be, by being awake and responsive to sensory input; objects cannot be conscious in the way mental states are. Still, having noninferential thoughts about states of one's liver presumably would not make those states conscious (Block, 1995b). But not every state can count as conscious. A state can be conscious only if being in it, even when the state is not conscious, results in one's being conscious *of* something, and states of the liver do not qualify (Rosenthal, 2000b).

Dretske has argued that a mental state's being conscious cannot consist in one's having a HOT about it, since there are cases in which a state is conscious without one's being conscious of it. Dretske offers the example of consciously seeing two scenes that differ in some single way without one's consciously noticing that they differ at all. Since one does not notice that the scenes differ, one also does not notice the difference between one's conscious visual experiences of the scenes. But every part of the two experiences is presumably conscious. So, if one is not conscious of that part of the experiences in respect of which they differ, that part is a conscious experience of which one is not conscious (Dretske, 1995). But all that Dretske's case shows is that one need not be conscious of that part *as* the part that makes a difference between the two experiences, not that one is not conscious of that part in some other way (Seager, 1999; Byrne, 1997; Rosenthal, 1999). It may well be that we are conscious of all our conscious experiences.

Conscious states presumably occur not just in humans, but in other animals as well. So perhaps conscious states occur even in animals whose mental functioning is too primitive to accommodate HOTs (Block, 1995a; Dretske, 1995; Byrne, 1997). Indeed, Carruthers (2000) actually argues that few if any nonhuman animals have HOTs. And he concludes that they lack conscious states, though many will resist that conclusion.

In any case, the reasons for thinking that few nonhuman animals have HOTs are not fully convincing. Carruthers (2000) argues that animals with HOTs would also have thoughts about the mental states of others. And he holds that having thoughts about the mental states of others would express itself in deceptive behavior, which he urges nonhuman animals do not engage in (cf. Povinelli, 1996). But it is arguable that many nonhuman animals do engage in deceptive behavior (Whiten, 1996; Whiten and Byrne, 1997). Nor, in any case, is it obvious that creatures would not have HOTs unless they had thoughts about the mental states of others (Ridge, 2001).

It might seem that nonhuman animals lack the conceptual resources needed to have HOTs. But HOTs do not require the elaborate conceptual apparatus characteristic of humans; they are simply thoughts that one is in states of particular types, states which we humans classify as mental.

At the same time, it is not obvious which nonhuman species do have mental states that are conscious. Though many such species plainly do sense and think, that does not show that their thinking and sensing are conscious; states can exhibit the characteristic causal roles of mental states without a creature's being conscious of being in those states. Some way independent of human subjective impressions is needed to establish which species do have mental states that are conscious.

As noted above, it is one thing for a creature to be conscious and another for its mental states to be conscious. Still, it might seem that a creature cannot be conscious unless at least some of its mental states are conscious; whenever humans are awake, after all, they are in some conscious states. But this may not hold generally. For a creature to be conscious it must function in characteristic mental ways, but it can do that without its mental states being conscious.

It is natural to think that a mental state's being conscious serves some useful function, such as enhancing the rationality of thinking and planning (Nelkin, 1996). But the function a mental state serves is a matter of its causal role, and the causal role a state has may well be largely unaffected by being accompanied by a HOT (Dretske, 1995).

Accompanying HOTs might, however, actually alter a state's causal role, and a state together with a

HOT will in any case have a different combined role from the state without any HOT. Indeed, it has been argued that HOTs enable the correction of plans that result from first-order processing (Rolls, 1998).

There are also experimental findings that subjects sometimes perform tasks better when stimuli are consciously perceived than when perceived nonconsciously (Merikle and Daneman, 1998). Since the relevant tasks require conscious thought, the difference may be due to operation of HOTs in the conscious cases.

There is a compelling intuition that our conscious states constitute some kind of unity, and one might object that a theory on which mental states are conscious in virtue of many distinct HOTs cannot do justice to that intuition (Shoemaker, in press). But since the content of each HOT is that one is, oneself, in some state, such reference to oneself will give rise to a conscious sense of unity (Rosenthal, in press c).

QUALITATIVE CONSCIOUSNESS

Perhaps the most important objection has to do with qualitative consciousness. It has been argued that HOTs cannot capture the enormous detail characteristic of conscious qualitative states (Byrne, 1997). And some have argued also that HOTs, which are nonqualitative, could not result in there being something it's like for one to be in various qualitative states (Byrne, 1997; Siewert, 1998; Balog, 2000).

It is easy to exaggerate the qualitative detail we are conscious of at any moment. It is well known that Parafoveal vision yields scant detail. More dramatically, recent work on change blindness shows that we often fail consciously to notice significant changes in a scene we are attentively looking at (Grimes, 1996; Rensink, 2000; Simons, 2000), which suggests that our impression of great conscious qualitative detail is erroneous (Dennett, 1991). And HOTs could presumably capture the detail present in any relatively small area of a sensory field on which one consciously focused.

We do not have concepts for all the individual qualities we are conscious of, but we have concepts for the ways those qualities vary. So HOTs can represent individual qualities comparatively. This may explain why we can judge whether qualities are the same far better when they are all present than when we must rely on memory (Raffman 1995). And, though concepts may be ill suited to capture the way qualitative states represent things, we are typically conscious of the relevant qualities in a way that lends itself to conceptualized description.

It is important not to place excessive demands on an explanation of qualitative consciousness. Very likely no explanation will reveal a conceptual or rational connection between nonconscious resources and conscious qualities (Levine, 2001), but scientific explanation seldom does that. Nor should we expect to discover an introspectible connection between conscious qualities and nonconscious resources, since nothing that is not conscious is available to introspection.

In any event, there is reason to think that HOTs do figure in there being something it's like to be in conscious qualitative states. We often come to be conscious of qualitative differences only when we come to have concepts fine-grained enough to draw those qualitative distinctions: for example as between similar musical instruments or tastes of wine. Such concepts would matter to how those experiences are conscious only if our thoughts about the experiences made a difference to how we are conscious of them (Rosenthal, in press a).

THE SCIENCE OF CONSCIOUSNESS

Although the foregoing arguments in support of the HOT hypothesis do not rely on empirical investigation, the hypothesis meshes fruitfully with scientific findings. Two examples already noted are change blindness and confabulated mental states. But there are others as well. As Weiskrantz (1997) has urged, a HOT model helps explain the phenomena of blindsight. Rolls (1998) argues for his linguistic version of the HOT hypothesis by appeal to different neural pathways that seem to subserve conscious and nonconscious stimulation. Dienes and Perner (2001) have appealed to the HOT model in distinguishing implicit from explicit knowledge and representation, and Dienes (in press) has applied the model in connection with implicit learning and subliminal perception.

The HOT model is particularly useful in explaining the finding by Libet (Libet, 1985) that the neural readiness potentials identified with subjects' decisions occur measurably in advance of subjects' awareness of these decisions, findings recently replicated and extended (Haggard and Eimer, 1999). It is natural to explain this result by supposing that the HOTs in virtue of which subjects become aware of their decisions occur measurably later than those decisions (Gomes, 1999; Rosenthal, in press d).

Frith and Frith (1999) report a number of studies in which functional brain imaging reveals neural activation in subjects who were asked to report

their mental states. Strikingly, conscious monitoring of states results in activation of a single brain area, medial frontal cortex, even when the states monitored are as disparate as pain, tickles, emotions aroused by pictures, and spontaneous thoughts. This activation does not occur cortically where the monitored states occur; so a single, independent brain mechanism seems to subserve the monitoring that makes possible the reporting of mental states. Since reports of one's mental states express one's thoughts about those states, it is inviting to construe that activation as indicating the occurrence of HOTs.

References

Armstrong DM (1978/1980) 'What is consciousness?'. *Proceedings of the Russellian Society* **3**(1978): 65–76; reprinted in expanded form in Armstrong, *The Nature of Mind*, St. Lucia, Queensland, Australia: University of Queensland Press, pp. 55–67, 1980.

Balog K (2000) Comments on David Rosenthal's 'consciousness, content, and metacognitive judgments'. *Consciousness and Cognition* **9**(2) Part 1: 215–219.

Block N (1986) Advertisement for a semantics for psychology. *Midwest Studies in Philosophy* **X**: 615–678.

Block N (1995a) 'On a confusion about a function of consciousness'. *The Behavioral and Brain Sciences* **18**(2): 227–247.

Block N (1995b) How many concepts of consciousness? *The Behavioral and Brain Sciences* **18**(2): 272–287.

Brentano F (1874/1973) *Psychology from an Empirical Standpoint* edited by Kraus O, English edn edited by McAlister LL, translated by Rancurello AC, Terrell DB and McAlister LL. London, UK: Routledge & Kegan Paul, 1973.

Byrne A (1997) Some like it HOT: consciousness and higher-order thoughts. *Philosophical Studies* **86**(2): 103–129.

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

Carruthers P (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge, UK: Cambridge University Press.

Chalmers DJ (1996) *The Conscious Mind*: *In Search of a Fundamental Theory*. New York, NY: Oxford University Press.

Dennett DC (1987) *The Intentional Stance*. Cambridge, MA: MIT Press/Bradford Books.

Dennett DC (1991) *Consciousness Explained*. Boston: Little, Brown and Company.

Dienes Z and Perner J (2001) When knowledge is unconscious because of conscious knowledge and vice versa. In: Moore JD and Stenning K (eds) *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, pp. 255–260. Mahwah, NJ: Lawrence Erlbaum Associates.

Dretske F (1995) *Naturalizing the Mind*. Cambridge, MA: MIT Press/Bradford Books.

Frith CD and Frith U (1999) 'Interacting minds – a biological basis'. *Science* **286**(i5445): 1692ff.

Gennaro RJ (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order-Thought Theory of Consciousness*. Amsterdam and Philadelphia: John Benjamins.

Goldman AI (1993) Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition* **2**(4): 364–382.

Goldman AI (2000) Can science know when you're conscious? epistemological foundations of consciousness research. *Journal of Consciousness Studies* **7**(5): 3–22.

Gomes G (1999) Volition and the readiness potential. *Journal of Consciousness Studies* **6**(8–9): 59–76.

Grimes J (1996) On the failure to detect changes in scenes across Saccades. In: Akins K (ed.) *Perception*, pp. 89–110. New York, NY: Oxford University Press.

Güzeldere G (1995) Is consciousness the perception of what passes in one's own mind?. In: Metzinger T (ed.) *Conscious Experience*, pp. 335–357. Exeter: Imprint Academic. Reprinted in: Block N, Flanagan O and Güzeldere G (eds) *The Nature of Consciousness: Philosophical Debates*, pp. 789–805. Cambridge, MA: MIT Press/Bradford Books, 1997.

Haggard P and Eimer M (1999) On the relation between brain potentials and awareness of voluntary movements. *Experimental Brain Research* **126**(1): 128–133.

Holmes DS and Frost RO (1976) Effect of false autonomic feedback on self-reported anxiety, pain perception, and pulse rate. *Behavior Therapy* **7**(3): 330–334.

Kant I (1787/1998) *Critique of Pure Reason*, translated and edited by P Guyer and AW Wood. Cambridge, UK: Cambridge University Press, 1998.

Kobes BW (1996) Mental content and hot self-knowledge. *Philosophical Topics* **24**(1): 71–99.

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

Libet B (1985) 'Unconscious cerebral initiative and the role of conscious will in voluntary action'. *The Behavioral and Brain Sciences* **8**(4): 529–539.

Locke J (1700/1975) *An Essay Concerning Human Understanding*, edited from the 4th edn. by PH Nidditch. Oxford, UK: Clarendon Press.

Lycan W (1996) *Consciousness and Experience*. Cambridge, MA: MIT Press/Bradford Books.

Marcel AJ (1983a) Conscious and unconscious perception: experiments on visual masking and word recognition. *Cognitive Psychology* **15**: 197–237.

Marcel AJ (1983b) Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology* **15**: 238–300.

Mellor DH (1977–78) Conscious belief. *Proceedings of the Aristotelian Society*, New Series, LXXXVIII: 87–101.

Merikle PM, Smilek D and Eastwood JD (2001) Perception without awareness: perspectives from cognitive psychology. *Cognition* **79**(1–2): 115–134.

Merikle PM and Daneman M (1998) Psychological investigations of unconscious perception. *Journal of Consciousness Studies* **5**(1): 5–18.
Millikan RG (1984) *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press/Bradford Books.
Natsoulas T (1993) What is wrong with the appendage theory of consciousness? *Philosophical Psychology* **6**(2): 137–154.
Nelkin N (1996) *Consciousness and the Origins of Thought*. Cambridge, UK: Cambridge University Press.
Nisbett RE and Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. *Psychological Review* LXXXIV(3): 231–259.
Peacocke C (1992) *A Study of Concepts*. Cambridge, MA: MIT Press/Bradford Books.
Povinelli DJ (1996) Chimpanzee theory of mind?: the long road to strong inference. In: Carruthers P and Smith PK (eds) *Theories of Theories of Mind*, pp. 293–329. Cambridge, UK: Cambridge University Press, 1996.
Raffman D (1995) On the persistence of phenomenology. In: Metzinger T (ed.) *Conscious Experience*, pp. 293–308. Exeter: Imprint Academic.
Rensink RA (2000) The dynamic representation of scenes. *Visual Cognition* **7**(1/2/3): 17–42.
Rey G (2000) Role, not content: comments on David Rosenthal's 'Consciousness, content, and metacognitive judgments.' *Consciousness and Cognition* **9**(2): 224–230.
Ridge M (2001) Taking solipsism seriously: nonhuman animals and meta-cognitive theories of consciousness. *Philosophical Studies* **103**(3): 315–340.
Rolls ET (1998) *The Brain and Emotion*. Oxford, UK: Clarendon Press.
Rosenthal DM (1986) Two concepts of consciousness. *Philosophical Studies* XLIX(3): 329–359.
Rosenthal DM (1993) Thinking that one thinks. In: Davies M and Humphreys GW (eds) *Consciousness: Psychological and Philosophical Essays*, pp. 197–223. Oxford, UK: Basil Blackwell.
Rosenthal DM (1997) Perceptual and cognitive models of consciousness. *Journal of the American Psychoanalytic Association* **45**(3): 740–746.
Rosenthal DM (1999) Sensory quality and the relocation story. *Philosophical Topics* **26**(1 and 2): 321–350.
Rosenthal DM (2000a) Introspection and self-interpretation. *Philosophical Topics* **28**(2): 201–233.
Rosenthal DM (2000b) Metacognition and higher-order thoughts. *Consciousness and Cognition* **9**(2): 231–242.
Rosenthal DM (in press a) *Consciousness and Mind*. Oxford, UK: Clarendon Press, 2003.
Rosenthal DM (in press b) Explaining consciousness. In: Chalmers DJ (ed.) *Philosophy of Mind: Contemporary and Classical Readings*. New York, NY: Oxford University Press, 2002.
Rosenthal DM (in press c) Unity of consciousness and the self. *Proceedings of the Aristotelian Society* **103**(3) (2003).
Rosenthal DM (in press d) The timing of conscious states. *Consciousness and Cognition* **11**(2) (2002).
Seager W (1999) *Theories of Consciousness: An Introduction and Assessment*. London and New York: Routledge.
Searle JR (1992) *The Rediscovery of the Mind*. Cambridge, MA: MIT Press/Bradford Books.
Sellars W (1963) Empiricism and the philosophy of mind. In: *Science, Perception and Reality*, pp. 127–196. London, UK: Routledge & Kegan Paul.
Shoemaker S (forthcoming) Consciousness and co-consciousness. In: Cleeremans A (ed.) *The Unity of Consciousness: Binding, Integration, and Dissociation*. Oxford: Clarendon Press.
Siewert CP (1998) *The Significance of Consciousness*. Princeton: Princeton University Press.
Simons DJ (2000) Current approaches to change blindness. *Visual Cognition* **7**: 1–16.
Staats PS, Hekmat H and Staats AW (1998) Suggestion/placebo effects on pain: negative as well as positive. *Journal of Pain and Symptom Management* **15**(4): 235–243.
Weiskrantz L (1997) *Consciousness Lost and Found: A Neuropsychological Exploration*. Oxford, UK: Clarendon Press.
Whiten A (1996) When does smart behaviour-reading become mind-reading? In: Carruthers P and Smith PK (eds) *Theories of Theories of Mind*, pp. 277–292. Cambridge, UK: Cambridge University Press.
Whiten A and Byrne RW (1997) *Machiavellian Intelligence, II: Extensions and Evaluations*. Cambridge, UK: Cambridge University Press.

Further Reading

Armstrong DM (1968/1993) *A Materialist Theory of the Mind*. New York: Humanities Press; 2nd revised edn. London, UK: Routledge & Kegan Paul, 1993.
Carruthers P (1989) Brute experience. *The Journal of Philosophy* LXXXVI(5): 258–269.
Dienes Z and Perner J (2001) The metacognitive implications of the implicit–explicit distinction. In: Chambres P, Izaute M and Marescaux P-J (eds) *Metacognition: Process, Function, and Use*, pp. 241–268. Dordrecht, Germany: Kluwer.
Dretske F (1993) Conscious experience. *Mind* **102**(406): 263–283; reprinted in Dretske, *Perception, Knowledge, and Belief*, pp. 113–137. Cambridge, UK: Cambridge University Press, 2000.
Haggard P (1999) Perceived timing of self-initiated actions. In: Aschersleben G, Bachmann T and Musseler J (eds) *Cognitive Contributions to the Perception of Spatial and Temporal Events*, pp. 215–231. Amsterdam, Netherlands: Elsevier.
Kobes BW (1995) Telic higher-order thoughts and Moore's paradox. *Philosophical Perspectives* **9**: 291–312.
Levine J (1993) On leaving our what it's like. In: Davies M and Humphreys GW (eds) *Consciousness: Psychological and Philosophical Essays*, pp. 121–136. Oxford, UK: Basil Blackwell.
Libet B, Gleason CA, Wright EW and Pearl DK (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness potential). *Brain* **106**(Part III): 623–642.

Lurz RW (in press) Advancing the debate between HOT and FO theories of consciousness. *Journal of Philosophical Research* **28** (2003).
Mellor DH (1980) Consciousness and degrees of belief. In: Mellor DH (ed.) *Prospects for Pragmatism*, pp. 139–173 Cambridge, UK: Cambridge University Press.
Perner J and Dienes Z (in press) Developmental aspects of consciousness: How much theory of mind do you need to be consciously aware? *Consciousness and Cognition*.
Rensink RA (2000) Seeing, sensing, and scrutinizing. *Vision Research* **40**(10–12): 1469–1487.
Rosenthal DM (2000) Consciousness and metacognition. In: Sperber D (ed.) *Metarepresentation: A Multidisciplinary Perspective*, pp. 265–295. New York, NY: Oxford University Press.
Rosenthal DM (2000) Content, interpretation, and consciousness. In: Ross D, Brook A and Thompson DL (eds) *Dennett's Philosophy: A Comprehensive Assessment*, pp. 287–308. Cambridge, MA: MIT Press/Bradford Books.
Rosenthal DM (in press e) Why are verbally expressed thoughts conscious?
Seager W (1994) Dretske on HOT theories of consciousness. *Analysis* **54**(1): 270–276.
Weiskrantz L (1986) *Blindsight: A Case Study and Implications*. Oxford, UK: Clarendon Press.
White PA (1988) Knowing more than we can tell: 'Introspective access' and causal report accuracy 10 years later. *British Journal of Psychology* **79**(1): 13–45.

Consciousness and Representationalism

Intermediate article

Benj Hellie, Sage School of Philosophy, Cornell University, Ithaca, New York, USA

CONTENTS

The representationalist theory of consciousness is the view that consciousness reduces to mental representation. This view comes in several variants which must explain introspective awareness of conscious mental states.

INTRODUCTION

Some mental states and processes are like something to their subjects; others are not. For instance, the states of seeing a stop sign, of hearing a screech, and of smelling gasoline are like something; as are the states of feeling fear, elation, or pain; as is the process of talking oneself through a problem. In contrast, states and processes that are not like anything to their subjects are accepted by both scientific and common sense psychology. Chomskian linguistic theories and Marrian theories of vision posit complex subpersonal operations, which make a difference to what one's mind is like to one only by their effects; common sense recognizes states of believing and intending that persist through dreamless sleep. States and processes that are like something to their subjects are conscious; otherwise not.

Among conscious states, what they are like to their subjects can differ: what seeing a red thing is like is standardly different from what seeing a green thing is like; what both are like differs from what smelling gasoline is like. A state has a 'phenomenal character' just in case it is conscious, or like something to its subject; two states have the same phenomenal character just in case what one is like to its subject is the same as what the other is like to *its* subject.

Phenomenal characters pose special problems for a fully naturalistic theory of the mind, for it may seem baffling how these properties can arise ultimately from interactions of particles and fields, or from processes in the brain. Wittgenstein famously wondered how *this* – his then current headache – could be a brain state; such bafflement is a proper reaction to the great difference in the ways in which phenomenal characters present