

Consciousness, Content, and Metacognitive Judgments

David M. Rosenthal

*City University of New York, Graduate School, Philosophy and Cognitive Science,
365 Fifth Avenue, New York, New York 10016-4309*

E-mail: dro@rucss.rutgers.edu

Because metacognition consists in our having mental access to our cognitive states and mental states are conscious only when we are conscious of them in some suitable way, metacognition and consciousness shed important theoretical light on one another. Thus, our having metacognitive access to information carried by states that are not conscious helps confirm the hypothesis that a mental state's being conscious consists in having a noninferential higher-order thought about that state. This higher-order-thought hypothesis readily explains the appearance to consciousness of confabulatory mental states—states that do not actually occur. This fits well with, and helps refine, the “No-Magic Hypothesis” advanced by Nelson and Narens (1990). © 2000 Academic Press

I. CONSCIOUSNESS AND METACOGNITION

Although our mental lives involve many conscious states, it is generally recognized that many mental states also occur without being conscious. The difference has to do with whether we are, in some way or other, conscious of the states in question. If one is in no way whatever conscious of a mental state, we do not count that state as being conscious, at least not as we intuitively draw the commonsense distinction between conscious and nonconscious states. It is another question just how we are conscious of those states we do count as conscious, a question that has often divided theorists and to which I will shortly turn. But if one is not at all conscious of a state, that state is not a conscious state.

Metacognition is, roughly, the access we have to whether, or how likely it is that, we know something. When we make judgments about whether we know something or how easily we will learn some item or even whether we have successfully learned it, these are metacognitive judgments. Such feeling-of-knowing (FOK) judgments, ease of learning (EOL) judgments, and judgments of learning (JOL) not only inform us about our cognitive condition (see, e.g., Forrest-Presley, Mackinnon, & Waller 1985; Flavell, Green, & Flavell 1995; Metcalfe & Shimamura 1994; Nelson 1992; Weinert & Kluwe 1987). They also, as T. O. Nelson (e.g., 1996 and Nelson & Narens 1990 and 1994) and his colleagues have usefully stressed, affect decision and action. Metacognitive judgments may also affect subsequent thought processes. The effects of metacognitive judgments on decision, action, and thought, moreover, go beyond the effects that the states we metacognitively judge ourselves to be in would have by themselves. I will say something more about this in Section IV.

Since metacognition consists in the mental access we have to our cognitive states

This article is part of a special issue of this journal on Metacognition and Consciousness, with Thomas O. Nelson and Georges Rey as Guest Editors.

and a mental state is conscious if one is, in some suitable way, conscious of that state, there is an obvious connection between consciousness and metacognition. Both involve some higher-order access to mental states we are in or seem to be in. It is an interesting question, about which I will say something later, whether this access can in either case be about mental states we are not actually in.

Despite their both involving access to our mental states, consciousness and metacognition are distinct phenomena. For one thing, we often make judgments about whether we have learned something or whether we know something without consciously accessing the relevant information—sometimes, even, when that information is, for some possibly transient reason, not currently accessible. Judgments of learning and feeling-of-knowing judgments often occur without current recall. We can, it seems, be conscious of knowing something without being conscious of the thing we know.

An especially dramatic example is the vivid feeling, familiar to us all, that a word or other piece of information is on the tip of one's tongue. In these situations, one has a conscious sense that the relevant informational state is there somehow, even though the state itself is not a conscious state.¹ We cannot in general explain this by saying that the information is there potentially, but not actually. That does happen, of course, when the information depends on putting together pieces of information already available. But if it is Mark Twain's real name that is on the tip of one's tongue, that information is not due to inference; so the information must somehow be actually present, though yet not currently conscious. Indeed, as William James usefully noted, even when we cannot consciously access anything about the word, the "gap" in our consciousness is often "intensely active" (James 1890/1950, p. 251); there is a vivid conscious difference between having one word on the tip of one's tongue and having another.

If consciousness and metacognition both trade on the higher-order cognitive access we have to our mental states, how can we explain the difference between them? We are in some way or other conscious of every conscious state we are in, but as I noted at the outset, it is a more difficult matter to specify the exact way in which we are conscious of our conscious states. The divergence between consciousness and metacognition that occurs in tip-of-the-tongue phenomena will help close in on the exact way in which we are conscious of our mental states when those states are conscious states.

When I have Mark Twain's real name on the tip of my tongue, I must be conscious of the particular state that carries that information. But I am not conscious of that state in respect of the specific information the state carries; rather, I am conscious of the state only *as* a state that carries that information. I am conscious that I know *what* Twain's real name is without, however, knowing *that* his real name is Samuel Clemens.

¹ The tip-of-the-tongue phenomenon that psychologists discuss involves having conscious access to partial information, perhaps the initial letters or phoneme of a name. In what follows I will use the rubric 'tip-of-the-tongue' in accord with its more commonsense usage to refer instead to cases in which we have a vivid sense, sometimes accurate, that the name or information could be accessed though we cannot get conscious access even to partial information.

This has implications for the way we must be conscious of a state when that state is conscious. For the state that carries this information to count intuitively as a conscious state, I must be conscious of that state in respect of the information it carries. It is not enough simply to be conscious of what question that information provides the answer to; for the state to be conscious, I must be conscious of what the information is. By contrast, tip-of-the-tongue experiences and, more generally, feeling-of-knowing judgments without current recall involve one's being conscious of informational states not in respect of the information itself, but only in respect of what questions that information would answer.

II. A THEORETICAL MODEL OF CONSCIOUSNESS

This explanation of the difference between a mental state's being conscious and our having metacognitive access to that state places constraints on acceptable theoretical models of consciousness.

A mental state is conscious only if one is conscious *of* that state. This is not circular; we understand what it is to be conscious *of* things independently of understanding what it is for mental states to be conscious. We are conscious *of* something when we see it or hear it or sense it in some other way. And we are also conscious of something when we have a thought about it as being present. I may now be conscious of somebody in the audience because I sense that person visually, but I may instead be conscious of that person only because I have a suitable thought about the person. We know of no other way of being conscious of things, and positing some third way would be theoretically idle without some independent grasp of what such a third way might consist in.

In any case, theorists have typically invoked the first way—being conscious of things by sensing them—in trying to explain what it is for mental states to be conscious. On this model, we are conscious *of* our conscious states by way of some kind of *inner sense*. This long-standing tradition, exemplified by philosophers otherwise as divergent as Locke and Kant as well as by many psychologists, has had a powerful and lasting influence on theoretical thinking about consciousness.

Nonetheless, there are compelling reasons to reject a sensory model of the way we are conscious of our conscious states. One reason has to do with sensory qualities. Sensing always involves a range of sensory qualities special to a particular modality. But intentional states are typically conscious independently of the occurrence of any sensory quality. And the sensory qualities that figure in our being conscious of sensory states are the qualities of those states themselves, not qualities of some higher order consciousness of the states.

These considerations are decisive against a higher-order sensing model of the way we are conscious of our conscious states. But the only other way we are ever conscious of things is by having thoughts about them. So it must be that the way we are conscious of our conscious states is by having thoughts about those states.

The earlier discussion of consciousness and metacognition reinforces this conclusion. When I have George Orwell's real name on the tip of my tongue but cannot recall it, I am conscious that I am in some state that carries the desired information, but I am not conscious of the state *in respect of* that information itself. How is this possible?

It cannot be that I simply *sense* the state, since sensing by itself does not differentiate things in the right way. We do, of course, sense things in respect of different aspects. I may see one side of a thing rather than another or hear it without seeing it or feel only one part of it. But sensing itself involves no intentional content, and typically occurs independently of any. So sensing cannot capture the difference between being aware of a state in respect of the specific information it carries and being aware of the state in respect only of some question that the information would answer.

But whether or not a state is conscious is responsive to this difference. So the way we are conscious *of* our conscious states must allow for such sensitivity. Since only intentional content can capture this difference, the higher-order state in virtue of which we are conscious *of* our conscious states must use intentional content to represent those states. So it must be that we are conscious of our conscious states by having suitable thoughts about them.

These thoughts will represent the states they are about in respect of the informational content those states have or, in the case of sensory states, in respect of their sensory quality. They will be thoughts to the effect that one is, oneself, in a state of a particular sort, where the relevant sort of state is ordinarily characterized in terms of an attitude held toward some intentional content or a particular sensory quality. Since the thoughts in virtue of which our states are sometimes conscious are about those states, I refer to them for convenience as higher-order thoughts (HOTs) (Rosenthal, 1986; 1990; 1993a; 1993b; 1997; 1998).

III. A FEW OBJECTIONS

I turn now to several apparent difficulties that might be raised in connection with this HOT model of consciousness. When I walk some place thoroughly engrossed in conversation, I must be conscious in some way or other of the obstacles I avoid. Indeed, I must have thoughts about these obstacles, since sensing would not by itself represent them in the way required to guide behavior. But, because I pay my environment no conscious heed, I am unaware of having any thoughts about the obstacles. I can be conscious of things without being conscious that I am.

Indeed, this is just what usually happens when I am conscious of my conscious mental states; the HOTs I have about those states are seldom, themselves, conscious thoughts. This is like the metacognitive access we have to our cognitive states. Sometimes we are aware of having such metacognitive access. But we may also be wholly unaware of having that access, as with judgments of learning that affect decisions without becoming conscious. In such cases the metacognitive access we have to our cognitive states is not conscious access.

The observation that HOTs are themselves seldom conscious takes care of two possible objections. There is no regress in explaining a state's being conscious by reference to one's having HOTs about it, since the top member of the hierarchy needn't be conscious. Indeed, it is an open empirical question just how far such a sequence of HOTs could run, as Nelson (1997) also notes in connection with the metacognitive hierarchy.

The second objection is that we are seldom aware of the HOTs this model posits. But we would be aware of these HOTs only if they were themselves conscious

thoughts, and on the current model they need not be. Distinguishing conscious from nonconscious HOTs also allows the model to explain the important difference between our being introspectively conscious of a state and our being conscious of it in the ordinary, unreflective way that occurs in the case of most of our conscious states. Since introspection involves being deliberately and attentively conscious of our conscious states, it requires that the HOTs in virtue of which we are conscious of those states are themselves conscious thoughts.

Not every mental state we are conscious of is a conscious state. The state must of course be one's own. But not every way of being conscious even of our own states makes those states conscious. I may be conscious of being in a state by applying a theory to myself or because somebody whose judgment I trust tells me. We can, however, readily rule out such counterexamples by positing that a HOT results in the target state's being conscious only when the HOT is not based on any conscious inference—that is, not based on any inference *of which* one is conscious. This does not mean that a state's being conscious hinges on a HOT's having some particular aetiology, as Alex Byrne (1997, p. 123) claims that the model implies. It is only *apparent* aetiology that counts; if it subjectively seems that we're conscious of a state *only* by inference, that state is not a conscious state.²

It has sometimes been held that we need not have occurrent HOTs about our conscious states, but need only be disposed to have such HOTs (Carruthers 1996; see Dennett 1991, ch. 10). But that will not do. Conscious states are those we are actually conscious of, not just potentially conscious of. And simply being disposed to have a thought about something does not make one conscious of that thing.

Having HOTs does not require having any particularly strong or elaborate concept of the self. For a HOT to attribute its target state to oneself, no more is needed than a conceptual distinction between oneself and everything else, a distinction that presumably any mammal can draw. Nor is any concept of mind required; HOTs need not characterize their targets as *mental*, but only as states. More generally, the concepts that figure in HOTs need not be anything like as rich as those which figure in typical human thinking. So the HOT model can apply not only in the human case but across the board to whatever other creatures may have mental states that we would count as conscious.

The higher-order states in virtue of which we are conscious *of* our conscious states cannot be higher-order sensations of target conscious states, since we are never aware of any qualities of such higher order states. But one might argue that we would not be conscious of whatever higher-order qualities do occur, since those higher-order states are themselves seldom conscious. Nonetheless, even when we are introspectively conscious of mental states and, hence, conscious of the relevant higher-order states, no higher-order qualities seem to occur.

IV. IMPLICATIONS AND SPECULATIONS

As we have seen, I can be metacognitively aware of a state that is not itself a conscious state if I am aware of that state but not of its specific informational content.

² For more on this, see Weisberg (unpublished).

I argued earlier that this confirms the HOT model of consciousness, since this difference cannot be captured by sensing our conscious states, but only by representing them in intentional terms.

It might be urged that this argument applies only to conscious states that are themselves intentional states, and that a different result might hold for conscious sensations. But the same thing applies to sensory states as well. I may feel that the color of some particular object is somehow on the tip of my imagistic tongue and, though I cannot at all now recall that color I make a feeling-of-knowing judgment that I would recognize it. But, since I cannot now recall the color, the color image I feel to be just out of reach is not a conscious image. This is possible only if a sensation's being conscious is a function of whether I represent the sensation merely as the image of whatever color the object has or as having a specific color. And this difference requires representing the sensation in intentional terms.

A state can be conscious, of course, without one's being conscious of every mental aspect of the state; introspective concentration often reveals mental aspects of conscious states that we had not been aware of. Indeed, it is plausible that we are seldom conscious of every mental aspect of our conscious states. This does not affect the contrast between conscious states and the kind of tip-of-the-tongue experience on which I have been focusing. Such experiences occur only when we are conscious of a state in respect of no mental aspect at all; instead, we are conscious only of some question that the state's informational content or qualitative character would answer. And if we were, after all, conscious of partial information, say the initial letter or phoneme of a word, then the informational state is to that extent a conscious state.

Similar remarks hold of the way we are conscious of the mental attitudes of our conscious intentional states. An intentional state will not be conscious unless one is conscious of oneself as holding some attitude toward the content in question, but one's conception of that attitude need not exhaust its mental nature. It is arguable, for example, that children three or younger, who fail the so-called false-belief task (see Perner, Leekam, & Wimmer, 1987; Flavell, 1988; and Wellman, 1990), have a conception of belief on which the content of beliefs match what is actually the case (see Rosenthal, 2000, Sect. IV). HOTs based on such a conception of believing would still suffice to result in the target states' being conscious.

Nelson and colleagues have proposed a model on which metacognition involves a level of psychological processing separate from the "object-level" processing that metacognitive monitoring is about. This raises the question of where in the brain such metalevel processing occurs; Nelson's speculation (1996, pp. 109–110, 1997) is that the frontal lobe plays a crucial role.

Such a modular picture of metacognitive processing is implicit in an analogy Nelson draws with levels of law courts, in which "an appellate court is at a meta-level relative to a superior court but at an object-level relative to a supreme court" (1997, p. 106). Courts do function, in effect, as distinct modules. But the notion of object- and meta-levels which Nelson (1996) borrows from Tarski (1956) does not, as Nelson assumes, imply that "the metalevel is in some sense separable from the object-level it refers to" (1996, p. 105). Metacognitive states could equally well occur interspersed, both functionally and neurophysiologically, with the cognitive states they are about. States belong to levels based solely on their intentional content; a state

gets sorted into the meta-level with respect to some other level just in case the state is about another state that belongs to that other level. No additional separation need occur.

A primary goal in the psychological study of metacognition is to discover the mechanisms that explain metacognitive access. This helps explain the appeal of a modular model, since modules might suggest such mechanisms. Nonetheless, it need not turn out that the relevant mechanisms are modular. Metacognitive access to our cognitive states is not, of course, all of one single type. Feeling-of-knowing judgments, judgments of learning, and ease-of-learning judgments may well rely to some extent on distinct mechanisms. And it may well be that the mechanisms that subserve one or another type of metacognitive judgment rely, in turn, on the functioning of relatively independent modules. But, by itself, a judgment's being metacognitive gives us no reason to think that making that judgment is subserved by any such independent module.

Nelson and Louis Narens (1990, p. 140) raise two central questions about metacognitive processing, both about mechanisms. One is the question of "[h]ow subjects consciously and validly monitor . . . nonrecallable" information they have. Also, Nelson and Narens have found, strikingly, that subjects take less time to monitor information they cannot currently recall than would be needed for a search of memory. This leads them to ask how subjects can "know about the presence/absence of an item in memory without first completing the memory search for it."

It will not help with these questions simply to note that metacognitive monitoring is, notoriously, less than fully accurate, since we still must explain the degree to which it does get things right. Instead, Nelson and Narens propose what they call the "No-Magic Hypothesis," on which feeling-of-knowing judgments do "not reflect any monitoring of unconscious information at all." Instead of "directly monitor[ing] a given unrecalled item in memory," such judgments "utilize only suprathreshold information about remembered attributes of the item (including incorrectly remembered information!), along with rules for how to utilize that information in the FOK [feeling-of-knowing] judgments" (1990, p. 158).

This important and suggestive hypothesis doubtless points toward a correct explanation of our metacognitive access of items currently unavailable for recall. Still, I want to register a qualification about how that hypothesis should be described. It is not strictly correct to say that feeling-of-knowing judgments "[do] not reflect any monitoring of unconscious information." Availability of information for recall is not the same as the the relevant informational state's being conscious. Indeed, all recall involves access to information that was not at the time already conscious. Rather, the No-Magic Hypothesis must hold simply that metacognitive access to an informational state unavailable for current recall involves reconstructing the relevant information from other information that is currently available. But even that other information is not conscious prior to its being accessed. And, if the process of reconstructing is not conscious, as is typically the case, the information feeling-of-knowing judgments use in reconstructing may remain unconscious, as well. Making feeling-of-knowing judgments does involve the monitoring of some unconscious informational states.

It is useful to ask how the No-Magic Hypothesis applies to tip-of-the-tongue phenomena, which involve metacognitive access to information unavailable for current

recall. I might be unable to recall Orwell's real name, but recall having once learned it or that I had previously had conscious access to it. Such a reconstructive inference would lead to my strong sense that the relevant informational state is there but beyond current recall.

This application of the No-Magic Hypothesis squares with the foregoing explanation of such cases, on which we can be conscious of an informational state that is not conscious because we are not conscious of the state in respect of its specific informational content. But it might seem as though the No-Magic Hypothesis actually points toward an even better explanation. On that hypothesis, we are conscious of the relevant informational state only by way of an inference from other information, and a state is conscious on the HOT model only if one is conscious of that state independently of any inference. But that is too quick. The HOT model does not preclude a state's being conscious if we are conscious of it by way of inference, but only if we are conscious of it by way of some *conscious* inference—an inference, that is, *of which* we are conscious. The inference the No-Magic Hypothesis posits for tip-of-the-tongue phenomena will seldom if ever be conscious.

V. CONTENT AND CONFABULATION

When metacognitive judgments rely on inference or plausible reconstructing of the relevant information, this may affect the actual content of the judgment itself. In tip-of-the-tongue cases, for example, the content of one's metacognitive judgment might be that one knows some item one cannot now recall. But the content of one's judgment might, instead, be that one once knew that thing and that one finds it difficult to believe that one has altogether forgotten it.

The difference in this case may well seem insignificant, but consider another example. Nelson reports that when subjects have answered questions of general factual knowledge, their immediate retrospective confidence judgments consistently overestimate the accuracy of those answers, though delayed retrospective confidence judgments do not exaggerate accuracy (1997, p. 108). That raises a question about just what the content is of subjects' immediate retrospective confidence judgments. Are they judging their actual confidence about the answer? Or are they instead simply judging that they have no other idea about what the correct answer is, or even that it does not seem to them that any confabulation figured in coming up with the previous answer?

One cannot settle the content of subjects' metacognitive judgments simply by asking them. Subjects often are not reliable about the content of their judgments, since the content of a state does not always match the way we are conscious of the state. It is well-known that we sometimes confabulate beliefs or desires to rationalize a certain situation or accommodate preconceived ideas (Nisbett & Wilson 1977; White 1988; and Wilson, Hodges, & LaFleur 1995). When we do, we are conscious of ourselves as having a belief or desire that we do not actually have.

Even sensory states can seem different, from a first-person point of view, from their actual mental character. This is easiest to see when the qualitative character of a sensation outstrips the way we are aware of it. One may be aware of a throbbing pain only as painful, and not also in respect of its throbbing quality, or aware of a

sensation of red not in respect of its particular shade, but only as red. But it also happens that one is conscious of a sensation in respect of a quality the sensation does not have at all. Dental patients are sometimes conscious of fear and the feeling of vibration from the drill as though those sensations were pain, despite thorough local anesthetic or even the complete absence of the relevant nerve.

There is every reason to think that this happens with metacognitive judgments, as well. Even when it seems to me that I am judging one thing—say, my actual degree of confidence in something—it may be that I am in fact judging another—say, whether I have any better ideas about the matter at hand. Moreover, when I do metacognitively judge that I have a certain degree of confidence, that judgment, if it is subjectively noninferential, will lead to its seeming to me from a first-person point of view that I have the degree of confidence my judgment represents me as having. Mistaken metacognitive judgments can result in confabulated conscious states. Indeed, on Nelson and Narens's No-Magic Hypothesis, this happens very often.

Confabulated conscious states are states we are conscious of ourselves as being in even though the states do not actually occur. We are, in this way, actually conscious of states we are not in, but subjectively seem to us to belong to our stream of consciousness. The way we are conscious *of* our conscious states is therefore not factive. Being conscious of a state does not imply that the state exists, nor if it does that its mental properties match the way we are conscious of it.

The occurrence of confabulated conscious states is important for squaring the No-Magic Hypothesis with the claim made earlier about tip-of-the-tongue experiences. Such an experience occurs, I argued, if one does not consciously recall anything about Orwell's real name, but one is nonetheless conscious of being in a state that carries the relevant information. But on the No-Magic Hypothesis, the name may seem subjectively to be available to me only because I have access to my memory of having learned what his real name is and access to my belief that I have not forgotten it. I need not, on the No-Magic Hypothesis, have access to any actual state carrying the information that the name is Eric Blair. How can that be reconciled with the idea that I am, after all, conscious of a state that carries the relevant information?

A state's being conscious is a matter of how our states seem to us subjectively, from a first-person point of view. Having metacognitive access to a state sometimes has such a subjective component, but it is also partly a matter of what influence actual informational states have on our cognitive processes. So it may well be that, although it seems subjectively that I am conscious of being in some particular informational state, that appearance of being in that state is due simply to confabulation. And according to the the No-Magic Hypothesis, that is what may happen in the case under consideration. It is subjectively just as though I have access to an informational state carrying the relevant name without having access to the information that state carries. But what I actually have access to may only be some reconstruction of such a state from my remembering having learned the relevant information.

It will be useful to return to the effect metacognitive judgments have on action, decision, and thought processes. When the state we make metacognitive judgments about is confabulatory and does not actually occur, plainly it can have no causal influence on action, decision, or further thought processes. Whatever causal impact occurs in these cases must be due to the metacognitive judgment alone. Moreover,

states with distinct content presumably have different causal properties. So even when a metacognitive judgment represents an actually occurring state as having content or quality different from the state's actual mental properties, the judgment and the target state will have different causal powers. Metacognitive judgments may well often affect things in ways that go beyond the causal properties of the states we metacognitively judge ourselves to be in.

Similar remarks hold about a mental state's being conscious. If a conscious state is wholly confabulatory, any causal impact will be due to the HOT that represents us as being in the confabulatory state. And, if a HOT represents an actually occurring state as having different mental properties from those it actually has, the HOT and target state will correspondingly diverge in causal influence. Indeed, even when a HOT or a metacognitive judgment accurately represents an occurring target state, the HOT or metacognitive judgment is still a distinct state from its target, with distinct mental properties. Unlike a target state, a HOT or metacognitive judgment has content that represents one as being in some other cognitive or mental state. So the causal properties of the HOT or metacognitive judgment will differ from those of the target state.

Nelson and his colleagues (e.g., 1996 and Nelson & Narens, 1990; 1994) have stressed the causal role of metacognitive judgments concerning ease and success of learning that on such things as allocating study time. One spends less if any time studying things one takes oneself to know. But it's worth noting that this is a somewhat specialized context, and that apart from questions about how much to study metacognitive judgments may play little useful role in influencing action, decision, and other thought processes.

It is often held that a mental state's being conscious also plays an important role in influencing action, decision, and thought processes. Only when a state is conscious, it is urged, can we use the state in rational planning. This common assumption idea is, however, a lot less plausible than is usually recognized. A mental state's influence on rational processes must in some way be a function of that state's intentional content, and the state's content is the same whether or not the state is conscious.

It may seem tempting to assign special efficacy to a state's being conscious because it may seem that a state's playing a rational role in decision, thought, and action is a matter of an agent's consciously making use of that state or its informational content. And how could an agent use information that was not conscious? But once we put aside metaphorical ideas about free will, it is plain that a subject's using the informational content of its intentional states must somehow depend on the causal impact of those intentional states. And that again raises the question why the causal properties of states with particular content would not suffice. What relevant causal influence does a subject's being conscious of that content add? Indeed, findings by Benjamin Libet et al. (1983) (See also Libet, 1985), replicated and extended by Patrick Haggard and colleagues (Haggard & Eimer, in press; Haggard, Newman, & Magno, 1999), confirm that our subjective awareness of decisions occurs measurably later than the actual events of deciding. It is the deciding that matters, not our consciousness of it. Related considerations have been advanced by Daniel Wegner and Thalia Wheatley (1999).

The HOT model is particularly well-suited to explain cases in which the way a

conscious state appears to us differs from the way it actually is. Whatever the actual character of a mental state, that state, if conscious, is conscious in respect of whatever mental properties one's HOT represents the state as having. And, since HOTs make us conscious of ourselves *as* being in particular states, it is even likely that a HOT together with the target state it is about will be indistinguishable, subjectively, from that HOT's occurring in the absence of its target. In this way, we confabulate being in a state that we are not actually in. It is unlikely that any other explanatory model can accommodate these disparities between the character of our actual mental states and the states we subjectively seem to ourselves to be in.³

REFERENCES

- 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: Cambridge University Press.
- Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown and Company.
- Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris, and D. R. Olson, (Eds.), *Developing theories of the mind*. Cambridge: Cambridge University Press.
- Flavell, J. H., Green, F. L., and Flavell, E. R., (1995). Young children's knowledge about thinking. *Monographs of the Society for Research in Child Development*, **60**(1) (Serial no. 243), 1–95.
- Forrest-Presley, D. L., G. E. Mackinnon, and T. Gary Waller (Eds.) (1985). *Metacognition, cognition, and human performance*. Orlando, FL: Academic Press.
- Haggard, P., and Eimer, M. (in press). On the relation between brain potentials and awareness of voluntary movements. *Experimental Brain Research*.
- Haggard, P., Newman, C., and Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology*, **90**(Part 2, May), 291–303.
- James, W. (1890/1950). *The principles of psychology*. New York: Holt; New York: Dover.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences* **8**(4, December), 529–539, with open peer commentary: pp. 539–558, and Libet's reply, Theory and evidence relating cerebral processes to conscious will, pp. 558–566.
- Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (Readiness potential). *Brain*, **106**(Part III, September), 623–642.
- Metcalf, J., and Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press/Bradford Books.
- Nelson, T. O. (Ed.) (1992). *Metacognition: Core readings*. Boston: Allyn and Bacon.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, **51**(2), 102–116.
- Nelson, T. O. (1997). The meta-level versus object-level distinction (and other issues) in formulations of metacognition. *American Psychologist*, **52**(2), 179.
- Nelson, T. O., and Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), pp. 125–173. *The psychology of learning and motivation*. New York: Academic Press.

³ A slightly shorter version of this paper was presented on June 5, 1999, at a session on metacognition at the Third Annual Conference of the Association for the Scientific Study of Consciousness at the University of Western Ontario. I am grateful to Tom Nelson, my cosymposiast on that occasion, and to others at the conference for useful reactions.

- Nelson, T. O., and Narens, L. (1994). Why investigate metacognition? In J. Metcalfe and A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*, pp. 1–25. Cambridge, MA: MIT Press/Bradford Books.
- Nisbett, R. E., and Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* **LXXXIV**(3), 231–259.
- Perner, J., Leekam, S. R., and Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, **5**(2), 125–137.
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies*, **49**(3), 329–359.
- Rosenthal, D. M. (1990). *Why are verbally expressed thoughts conscious?* Report No. 32/1990, Center for Interdisciplinary Research (ZiF), University of Bielefeld.
- Rosenthal, D. M. (1993a). Thinking that one thinks. In M. Davies and G. W. Humphreys (Eds.), *Consciousness: Psychological and philosophical essays*, pp. 197–223. Oxford: Blackwell.
- Rosenthal, D. M. (1993b). State consciousness and transitive consciousness. *Consciousness and Cognition*, **2**(4), 355–363.
- Rosenthal, D. M. (1997). A theory of consciousness. In N. Block, O. Flanagan, and G. Güzeldere (Eds.), *The nature of consciousness: Philosophical debates*, pp. 729–753. Cambridge, MA: MIT Press.
- Rosenthal, D. M. (1998). Consciousness and its expression. *Midwest Studies in Philosophy*, **XXII**, 294–309.
- Rosenthal, D. M. (2000). Consciousness and metacognition. In D. Sperber (Ed.), *Metarepresentation: Proceedings of the tenth Vancouver Cognitive Science Conference*. New York: Oxford University Press.
- Rosenthal, D. M. (in preparation). *Consciousness and mind*. Oxford: Clarendon Press.
- Tarski, A. (1956). The concept of truth in formalized languages. In *Logic, Semantics, and Metamathematics*, pp. 152–178. Oxford: Clarendon Press.
- Wegner, D. M., and Wheatley, T. P. (1987). Apparent mental causation: Sources of the experience of will. *American Psychologist*, **54**(7), 480–492.
- Weinert, F. E., and Kluwe, R. H. (Eds.) (1987). *Metacognition, motivation, and understanding*. Hillsdale, NJ: Erlbaum.
- Weisberg, J. (unpublished). If you can't stand the heat, get out of the kitchen: A HOT response to Byrne.
- Wellman, H. W. (1990). *The child's theory of the mind*. Cambridge, MA: MIT Press/Bradford Books.
- White, P. A. (1988). Knowing more than we can tell: 'Introspective access' and causal report accuracy 10 years later. *British Journal of Psychology*, **79**(1), 13–45.
- Wilson, T. D., Hodges, S. D., and LaFleur, S. J. (1995). Effects of introspecting about reasons: Inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, **69**(1), 16–28.

Received March 2, 2000