In *Consciousness*, by Rita Carter, London:
Weidenfeld & Nicolson, 2002, pp. 45-47;
Also available as *Exploring Consciousness*,
Berkeley and Los Angeles: University of
California Press, 2002.

THE HIGHER-ORDER MODEL OF CONSCIOUSNESS

All mental states, including thoughts, feelings, perceptions, and sensations, often occur consciously. But they all occur also without being conscious. So the first thing a theory of consciousness must do is explain the difference between thoughts, feelings, perceptions, and sensations that are conscious and those which are not.

At bottom, the difference stems from the fact that, when a mental state is conscious, we are conscious of being in that state. This is clear from considering that, when one isn't in any way conscious of having a thought, sensation, feeling, or perception, that state does not count as a conscious state. And it is arguable that the best explanation of how we are conscious of being in those states is that we have thoughts that we are in those states. Because these thoughts are about other mental states, I call them higher-order thoughts (HOTs).

When a mental state is conscious, we are conscious of it in a way to seems, subjectively, direct and unmediated. We can account for this by providing that the HOTs we have about our mental states do not rely on any conscious inference. Suppose I want something to eat. If I come to have a thought about that desire because I infer from something about my behavior that I want to eat, my desire won't be conscious. My HOT must arise independently of any such inference.

HOTs need not themselves be conscious. My HOT about my desire to eat won't itself be conscious unless I have a second HOT about it. This explains why we usually aren't conscious of having any HOTs.

Sometimes, however, we are conscious of our HOTs. Though mental states are usually conscious in a relatively unreflective, unfocused way, we sometimes deliberately attend to some particular

thought or feeling. We thereby become introspectively conscious of that state. In such cases, we introspectively focus on the state; we are conscious not only of the state, but of our being aware of the state. These are cases in which our HOTs are themselves conscious thoughts. Not only is my desire to eat conscious; my thought that I want something to eat is conscious as well.

It is sometimes urged by critics that the HOT model explains only introspective consciousness. This idea stems from the unfounded assumption that HOTs must themselves be conscious. But since HOTs typically aren't conscious, we can invoke them to also explain ordinary, nonintrospective consciousness as well as introspective consciousness.

Because HOTs are thoughts to the effect that one is in some particular state, it must make reference to oneself. So it might seem that nonhuman animals lack the conceptual resources needed to have HOTs. But thoughts that refer to oneself needn't make use of a sophisticated concept of the self. All that's required is a concept of the self strong enough to distinguish oneself from everything else. And it's clear that many nonhuman creatures must be able to frame such thoughts.

Still, some nonmental creatures presumably do lack the ability to have HOTs. But this isn't on the face of it a difficulty for the model. Many nonhuman species do of course function in ways that establish pretty firmly that they sense things and even have some simple thoughts. But such functioning does not by itself also establish that those perceptions and thoughts are themselves conscious. We can't just rely on subjective impressions to establish which nonhuman species do have conscious thoughts and sensations. When we do figure out which nonhuman species do have sensations and feelings that are conscious, it could easily turn out that all such species also have the mental resources needed for HOTs.

There is a difference between a mental state's being conscious and a creature's being conscious. A creature is conscious if it is awake and can receive sensory information. But that can readily happen even if the creature were never in any way aware of its mental states. So we can't infer from an animal's being conscious that its mental states are ever conscious. In the human situation,of course, being awake always goes with being in some mental states that are conscious, but that need not hold generally.

Because HOTs usually aren't conscious and people are

normally unaware of them, we cannot establish their occurrence by being conscious of them. HOTs are, rather, theoretical posits whose occurrence is established by theoretical considerations.

Indeed, the model meshes fruitfully with many scientific findings. It helps, for example, in explaining phenomena such as change blindness and blindsight. The HOT model readily accommodates the nonconscious sensing that occurs in blindsight. And the model can explain change blindness as due to the failure of sensations that result from changes in a scene to become conscious. The model also helps explain cases in which subjects confabulate having various thoughts and desires. These subjects have HOTs that they have such thoughts and desires, and these HOTs make it seem subjectively that they have those states even though they don't.

The HOT model is especially useful in explaining Libet's finding that the neural readiness potentials identified with subjects' decisions occur measurably after their awareness of those decisions. This is to be expected, since the HOTs in virtue of which subjects become aware of their decisions presumably occur measurably later than those decisions.

Some recent brain-imaging work indicates that, when subjects are asked to report their mental states, neural activation occurs in a single brain area, namely, medial frontal cortex. This is so even when the states monitored are themselves very different in kind, for example, pain, tickles, emotions aroused by pictures, and spontaneous thoughts. [For a summary, see Chris D. and Uta Frith, "Interacting Minds--A Biological Basis," *Science* 286, i5445 (November 26, 1999): 1692ff.] The location of activation due to monitoring is also distinct from the locations of the various types of monitored state. This suggests that a single, independent brain mechanism subserves the monitoring that enables us to report our mental states. Since reports of one's mental states express one's thoughts about those states, it is inviting to construe this neural activation as indicating the occurrence of HOTs.