Why Are Mental States Ever Conscious?

David Rosenthal
CUNY Graduate Center
Philosophy and Cognitive Science
http://davidrosenthal1.googlepages.com/

Taipei, 19 June 2008

OVERVIEW
[including 10 slides skipped at ASSC-12]

1. Problems and Theories
2. Utility and Evolution
3. Why Sensing Is Ever Conscious
4. Why Thinking Is Ever Conscious
I. Problems and Theories

- We use the term ‘consciousness’ for several different phenomena, which it’s crucial to distinguish.
- The phenomenon I’ll focus on today is the consciousness of mental states—e.g., of thoughts, volitions, perceptions, and sensations—which I’ll call state consciousness.
- I won’t talk at all about a person or other creature’s being conscious—as against being, e.g., asleep or anaesthetized.

So I won’t be concerned today with how sleep differs from a wakeful state, as in Giulio Tononi’s work on the thalamocortical system, nor with vegetative wakefulness without awareness, as in work by Steven Laureys.

- My focus will instead be on the property in virtue of which some mental states, but by no means all, are conscious—i.e., the property in virtue of which conscious mental states differ from mental states that are not conscious.
- I’ll call that property state consciousness.
But even when we restrict our focus to the property of mental states’ being conscious, there are several different questions that theorists address.

One question has to do with the neuronal correlate of consciousness (NCC): Which neuronal events occur when, but only when, mental states are conscious—and so are presumably those neuronal events in virtue of which mental states are conscious.

But there is a pivotal distinction, sometimes ignored, that we must take account of in connection with the NCC.

Mental states occur not only consciously, but also nonconsciously, as in subliminal perception and masked priming (e.g., Marcel 1983a, b; Breitmeyer and Ögmen 2006).

So we must distinguish between the state itself—which makes us conscious of something (what I'll call transitive consciousness)—and that state itself being conscious (state consciousness).

So there are two NCCs—one for the state, independent of its being conscious, and a second for the state’s being conscious (cf. Lau & Passingham 2006; Lau 2008 [Block 2005, 2007]).
I’ll come back shortly to this distinction between state consciousness and transitive consciousness.

But first I want to mention a second question, wholly independent of the NCC, that some theories focus on, namely: What it is for a mental state to be conscious.

There are a number of such theories in the literature, and they are in a way prior to theories about the NCC, since we can investigate the NCC of state consciousness only if we understand what state consciousness is.
Several theories focus on this second question—what it is for a state to be conscious: E.g., higher-order theories: A state is conscious if the subject is conscious of that state (e.g., Rosenthal 2005); first-order theories: A state is conscious if it makes one conscious of something (e.g., Dretske 1995, 1996); intrinsicalism: A state is conscious if it's self-directed (Brentano 1874); and global-workspace theory: A state is conscious if many other states can access its content (Baars 1988; Dehaene and Naccache 2001; Dehaene et al 1998 [Block 1995, 2007]).

Most theories that focus on the consciousness of mental states address one of these two questions: the NCC or what it is for a state to be conscious.

But my concern today is with a third question: Given that not all mental states are conscious, why is it that any are conscious?

One way to approach this third question would be in terms of neural processes: Perhaps there is something in the way the brain works that gives rise to the NCC (for state consciousness) for some states but not others, and consciousness hence results in those states' being conscious.
But I want instead to approach this third question at the psychological level, and ask what psychological processes, if any, can explain why some mental states do come to be conscious.

A standard answer is that some states are conscious because their being conscious is useful for the organism’s functioning. I’ll argue briefly in II that this line, however inviting, is not promising.

But first I want to ask whether the various theories about what consciousness is have any implications for the question why mental states are ever conscious.
On a first-order theory, a state is conscious if being in it makes one conscious of something. Since that has utility, we can expect many states to be conscious.

But subliminal perception, e.g., makes one conscious of things, even though those states aren’t conscious. So that line fails.

On intrinsicalism, a state is conscious if it’s self-directed in a way that makes one conscious of being in it. But that theory says nothing about why some states would be thus self-directed—and indeed suggests that mental states are all self-directed, and so all conscious.

Global-workspace theory does distinguish conscious from nonconscious states, and global access to a state implies a lot of utility for that state.

But I’ll argue in [II] that a state’s being conscious has little or no utility—and that speaks (indirectly) against global-workspace theory, as well as any other utility-based account of why some states are conscious.

In summary so far:
We can’t explain why some (but not all) mental states are conscious with a first-order or intrinsicalist theory, and perhaps not global-workspace theory.
So a higher-order theory ([III-IV]) may offer the best chance of explaining that.
II. Utility and Evolution

- On global-workspace theory, it’s useful for mental states to be conscious.
- So perhaps we can explain why some states are conscious by appeal to the utility of—and consequent evolutionary selection pressures for—some mental states’ being conscious.
- As we’ve seen, we must distinguish a state’s being conscious from that state’s making one conscious of something.

It’s of course useful for states to make one conscious of things; our question is what utility is added by those states’ also being conscious states.

I’ll first present a theoretical argument that there is little if any added utility, and then supplement that with some empirical and methodological considerations.

Here is the theoretical argument: The utility of being in any mental state, whether or not that state is conscious, will be mainly a function of that state’s representational character—i.e., what the state represents, and how.
That's plainly so for qualitative states, such as sensations and perceptions: Their utility stems from their representing the world reasonably accurately.

Perceptual and bodily sensations are useful only if the way they represent the environment and one's own body enables behavior conducive to one's well-being.

And the same holds for intentional states, such as thoughts and desires; they have utility if they represent the world as it is, and how one wants it to be.

Thus theories of intentional content insist that content in some way tracks reality.

So mental states—both qualitative and intentional—are useful because of their representational character. A state's being conscious, moreover, is independent of its representational character—since any mental state, whatever its representational character, can occur both consciously and not consciously.

So the utility of one's being in a state will itself be largely independent of its being conscious; and its being conscious will add little if any utility to what it has when not conscious.
A concrete example: Suppose I want a beer, and I also think that there’s beer in the refrigerator. Together, the desire and thought cause me to go get a beer from the fridge, with various perceptions guiding me to the fridge—and to the beer.

These states cause my behavior because the causal potential of these states reflects their representational content—and since such states can occur without being conscious, that content is independent of their being conscious.

So utility in causing behavior is as well.

Experimental work bears all this out.

Complex deliberations about consumer choices are better when they’re not conscious (Dijksterhuis et al 2006; cf. Bargh 2002).

Functionality is actually enhanced when deliberation is not conscious.

Also, work by Libet (1985), replicated and refined by Haggard (e.g., 1999), suggests that volitions occur initially without being conscious, and become conscious only after they’ve had their causal impact (cf. Soon, Brass, Heinze, and Dylan, 2008).

And a state’s becoming conscious can’t add utility once that causal process has started.
It’s often said that conscious monitoring is needed to rationally correct or adjust our thinking. But that’s doubtful.

We seldom consciously rehearse the steps when we do adjust our thinking— and when we do, it’s typically awkward and slow. Typically we seem subjectively just to come to see things more clearly.

The Dijksterhuis and related findings bear this out. Since deliberating about complex choices—both in and outside the lab—yields better results when not conscious, adjusting our thinking must be efficacious even when it isn’t conscious.

Some types of behavior (in humans, or even in general) may well occur only when the relevant volitions are conscious.

But that doesn’t by itself show that the consciousness of such volitions enables those behaviors.

Indeed, the Libet-Haggard-Soon findings suggest that volitions initiate behavior while still nonconscious, and become conscious only after doing so.

So even when the behavior occurs after the states become conscious, their being conscious may not cause the behaviors; rather, both the behavior and the states’ being conscious may have a single cause.
So much for intentional states; what about perceptions and other qualitative states? It seems subjectively as though subliminal perceiving has limited utility—that perceiving affects rational thinking only when the perceiving is conscious. But if that is so, it might be just because we perceive so much consciously that we’re used to recruiting only the conscious cases of perceiving. The bilaterally blindsight monkey, Helen, came to be able to catch flies and pick up currants. (A narrow upper area was spared, but seems not to have figured in this behavior.)

Attention occurs even when perceiving isn’t conscious (Koch and Tsuchiya 2007; Kentridge, Heywood, and Weiskrantz 2004; Lamme 2003), again suggesting that perceiving need not be conscious to play a useful, rational role. And if perceiving needn’t be conscious to induce attention, nonconscious perceiving is likely useful in many other ways as well. Like attention, the representational character of perceiving is independent of its being conscious. And since utility is due to representational character, that utility is independent of the perceiving’s being conscious.
Larry Jacoby’s exclusion task is often said to show that perceiving must be conscious to figure in intentional action. Subjects are visually presented with a word, say, ‘reason’, and asked to complete a word stem, say ‘rea--’, with any word other than the presented word. When a word is presented for 500 ms, subjects see it consciously and mainly succeed in following the instruction; when it’s presented for only 50 ms, they report seeing no word—but tend to complete the stem with the word that was presented.

So subjects intentionally exclude a word only when they consciously see it. But that’s arguably not because intentional action requires consciousness: Subjects are instructed not to complete the stem with a word they see. But when they don’t see it consciously, they aren’t conscious of seeing it—so they think they don’t see it. It’s not that perceiving must be conscious for intentional action. Rather, in the 50-ms case, subjects think that they don’t see the word (though unbeknownst to them, it does still prime their responses).
Jacoby (1991) takes such exclusion-task findings to show that perception enables control of action only when it's conscious. But as just noted, they don't show that. And the control test is arguably question-begging (Fu, Fu, and Dienes 2007).

Also, nonconscious perception affects performance in learning of sequences and artificial grammars—even though subjects take themselves to be simply guessing (Fu et al 2007 and Dienes, Altmann, Kwan, and Goode 1995; cf. Turk-Browne, Jungé, and Scholl 2005).

So it's unclear that perceiving's being conscious is useful in producing behavior.

Prosopagnosiac, amnesiac, and blindsight subjects perceive nonconsciously, but lack “negotiable knowledge” of things they nonconsciously detect (Weiskrantz 1997, ch. 7).

But this might be because subjects are used to perceiving consciously and, like Helen before she became adjusted to her condition, subjects are not accustomed to recruit nonconscious perceptions.

Also, not all mental processing is intact in these disorders. So that may be partially responsible for subjects’ failure to recruit such nonconscious perceiving as occurs.
Jeffrey Gray (2004) argues that, though conscious qualitative states arise too slowly to affect online behavior directly, their utility likely consists in enabling late error detection.

But it’s unclear why persistent qualitative states couldn’t do that even when they’re not conscious; nonconscious sensations could operate downstream to correct errors—and also cause one to be conscious of the states themselves.

Gray assumes that qualitative states’ being conscious must have some utility; but that begs the question at hand.

We often consciously regard mental states as having utility—but that doesn’t show that their being conscious itself has utility. We may see as useful even if being conscious of it has no additional utility.

Lack of utility does not mean that a mental state’s being conscious is causally inert—i.e., epiphenomenalism.

Doubtless there are causal consequences for mental states’ being conscious.

But those consequences may well be too small, varied, or lacking in utility and reproductive advantage to explain why some mental states come to be conscious.
In summary:
A mental state’s utility hinges on its causal potential, and that causal potential tracks representational content. And since each state’s representational content is independent of that state’s being conscious, so is each state’s utility. This independence is confirmed by various empirical findings.

So we can’t explain why some mental states come to be conscious by appeal to the utility of their being conscious—and so not by appeal to global-workspace theory, nor to evolutionary adaptive value.

III. Why Is Sensing Ever Conscious?

Since mental states occur consciously so often, it’s natural to assume that their being conscious adds some functionality. But theory and evidence point in the opposite direction. So we need some other explanation of why mental states are so often conscious. That’s what I’ll sketch in the rest of my talk (see Rosenthal 2005; 2008; and in preparation).
It must be that mental states often cause one to be aware of them—thereby causing themselves to be conscious.

States in their nonconscious form cause behavior and other mental states—and also cause an awareness of themselves.

So our problem is to explain why mental states cause that higher-order awareness.

As we saw: If this higher-order awareness is intrinsic to each state, it’s unclear why some states have it when others don’t; first-order theories can’t accommodate nonconscious states; and global-workspace theory wrongly ascribes utility to a state’s being conscious.

So a higher-order theory may be best able to explain why mental states are sometimes—but not always—conscious.

I’ll use the higher-order-thought (HOT) theory I’ve developed elsewhere, since the alternatives (higher-order perceiving and dispositions for HOTs) face serious problems.

On that theory, a state is conscious if an accompanying thought makes one aware of that state—and that thought relies on no conscious inference or self-observation (so that the awareness will be subjectively direct and “from the inside”).

What concepts do these HOTs require?
HOTs about qualitative states must conceptualize the states they are about in respect of their mental qualities.

How do we come to have such concepts?

States with mental qualities are states in virtue of which one can discriminate a range of perceptible properties.

States with a particular mental quality enable one to pick out the corresponding perceptible property from others in that family—e.g., the family of colors.

This suggests an account of what mental qualities are—and hence an account of our concepts of those qualities.

The physical colors—the color properties we perceive physical objects to have—form a quality space of similarities and differences that we can perceive among them.

Mental qualities enable us to discriminate the corresponding perceptible properties; so we can fix them by their location in a corresponding quality space: The mental quality red is the property whose position in its quality space corresponds to the location of perceptible physical red in its quality space. And that fixes the concept of the mental quality red.

Still: How would a creature ever come to have such concepts?
Perceiving something as red involves not only the mental quality of red, but also one’s conceptualizing the perceived object as being (physically) red.

So in perceiving something as red, we deploy the concept of (physical) red.

Sometimes, however, seeing is erroneous. And sometimes a creature comes to detect that its perceiving an object as being red, e.g., is erroneous.

And doing so means detecting that one has been in a state that corresponds in the relevant way to red objects—even though there was no relevant red object.

But detecting that one is in a state that corresponds in that way to red objects involves having a concept of the mental quality of red: the concept of a state that corresponds to physical red.

So coming to be able to detect erroneous perceiving results—just by itself—in the creature’s having a concept of the relevant mental quality.

Since these concepts extrapolate from concepts of perceptible properties, this will readily generalize.

The ability to detect perceptual error gives rise to the concepts needed for HOTs.
But error detection typically involves past perceptual states. How does a creature come to have HOTs about its current qualitative states—HOTs that result in one’s being aware of oneself as now being in those states? Detecting erroneous perceiving involves having thoughts about one’s perceptions—in respect of their mental qualities. So if one detects erroneous perceiving reasonably often, one will get habituated to having thoughts about one’s perceiving.

Any perception whatever has the potential to be erroneous. So a creature used to detecting perceptual error will be disposed (to some extent) to have thoughts about its perceptions. Also: Each perception conceptualizes the thing perceived as having some perceptible property. And concepts of mental qualities extrapolate from those of perceptible properties. So each perception will dispose one to have a HOT about that perception itself—a HOT that characterizes the perception in respect of the relevant mental qualities.
Given habituation to recognize error, these HOTs are prompted by the perceptions themselves. So there is no conscious inference from perceiving to HOT.

And self-observation doesn’t figure at all in the production of these HOTs. Since the resulting HOTs are independent of any conscious inference and independent of any conscious observation of oneself, the awareness those HOTs give us of our perceptions will be subjectively immediate and “from the inside.”

So perceptions and bodily sensations will often come to be conscious in any creature with a suitable ability to recognize its own erroneous perceiving.

**In summary:**
A creature that can detect perceptual error must have some concept of its relevant perceptual states. This error detection will habituate such a creature to have thoughts (HOTs) about its current perceptual states. Such HOTs will often rely on no conscious inference or self-observation; so they will result in the perceptions’ being conscious.
IV. Why Is Thinking Ever Conscious?

- Still, perceptual error correction doesn’t help with the consciousness of thoughts and volitions, which have intentional content but no qualitative character.
- As with qualitative states, explaining why some intentional states come to be conscious involves two questions:
  1. How do the relevant concepts arise?
  2. Why do the relevant HOTs ever occur?

Many creatures that can think doubtless have no concept of intentional content.
- And coming to have such concepts is far more difficult than coming to have concepts of mental qualities.
- Mental qualities correspond to properties that objects are perceived to have, which many creatures will have thoughts about.
- By contrast, the intentional content of thoughts corresponds to the semantic properties of speech acts.
- So one’s having the concept of intentional content very likely builds on having some concept of the meaning of speech.
How might such a concept arise?
Creatures that describe speech in semantic terms might come to posit, in a folk theoretical way, inner states that cause speech performances—i.e., thoughts (Sellars 1956, 1968, 1975). (Compare the parallel move from perceptible properties to mental qualities.)

This folk-theoretic extrapolation from semantic concepts would result in new concepts that characterize thoughts in terms of their intentional content.

Such creatures (ur-people) would thereby come to have the folk-psychological concept of a thought—i.e., the concept of an inner state with intentional content.

Why, then, would these concepts come ever (indeed, reasonably often) to figure in HOTs about these creatures’ own intentional states?
Simply positing inner causes of speech acts would lead our creatures to ascribe intentional states to themselves—inferring from their own observed verbal behavior.

These observation-based self-ascriptions of intentional states would express thoughts about those intentional states.

But being observation based, those self-ascriptions would be on a par with third-person ascriptions of states to others.
Because these ascriptions rely solely on self-observation, the thoughts they are about would not be conscious:

Consider ascribing a thought to yourself relying solely on self-observation.

But as our creatures gained facility at such self-ascription, they would no longer need to observe their own behavior.

Simply being disposed to say (and to do) various things would come in time to cause HOTs not based on any conscious self-observation.

How could that come to happen?

Our ur-people would in time come to self-ascribe fluently their own thoughts and desires—still relying solely on self-observation.

But with enough practice, simply having a thought would, by itself, dispose these creatures to have the very HOTs they had earlier arrived at by inferring from self-observation.

They would come to have HOTs about their own thoughts, independent of conscious inference and independent of self-observation.
These ur-people (our ancestors? [Sellars 1956]) will come to be disposed to self-ascribe thoughts as we now do—in a way that short circuits all self-observation and conscious inference.

The resulting HOTs would be subjectively direct and “from the inside,” since those HOTs would not rely on self-observation or conscious inference.

These creatures would be conscious of themselves as having various thoughts and desires. So those states would be conscious.

Having a thought often prompts a HOT that one has that (first-order) thought.

Such prompting is evident in the pragmatic equivalence of saying something and saying that one thinks that thing, i.e., of expressing a thought and describing oneself as having that thought.

It’s because our thoughts often prompt HOTs that we have those thoughts that (with minor qualifications that are irrelevant here) whenever we say something, we could as readily have said that we think that thing (Rosenthal 2005, ch. 10, esp. ?v).
Because HOTs are typically prompted by the very first-order thoughts they are about, they are typically accurate.

So this account explains not only why our thoughts and desires are often conscious, but also why our awareness of them is typically accurate.

One might object that because this account relies on linguistic ability, it cannot explain how the mental states of nonlinguistic creatures come to be conscious, as we assume they often do.

But that limitation has only to do with how intentional states come to be conscious.

Concepts for intentional states do very likely require language use—and indeed the ability to talk about such language use.

But concepts for mental qualities require only the ability to perceive, and to detect perceptual error.

So HOTs about qualitative states can arise in creatures with no linguistic ability at all.

And though it’s natural to assume that the sensations of nonlinguistic creatures are often conscious, there’s no reason to think that their thoughts ever are.
It's worth stressing that the issue about nonlinguistic creatures is not whether such creatures are in states that have intentional content; doubt they often are. The issue rather whether the intentional states of such creatures are ever conscious specifically in respect of their intentional content. States, such as perceptions, that have both qualitative character and intentional content might well be conscious—but only as qualitative states. Nonlinguistic creatures might never be conscious of themselves as being in states that have intentional content; nor does pretheoretic intuition say otherwise.

Summarizing this section: Concepts of states with intentional content arise, in creatures that can think about their own speech acts, by a folk-theoretic extrapolation from concepts for the semantic properties of speech acts. HOTs about one's own thoughts come to occur, then, when ascribing thoughts to oneself comes to be relatively automatic. Thoughts would come to conscious only in language-using creatures. But it’s likely that only the perceptual (qualitative) states of other creatures are conscious—and they’re conscious independent of language use.
We can’t explain why some states are conscious but not others by appeal to first-order or intrinsicalist theories, or to the utility of states’ being conscious. But the appeal to HOTs can explain—by way of error detection and the way we come to describe intentional states—why HOTs sometimes occur, and hence why the states they are about are conscious.