Mental Qualities and Consciousness

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

Efron Symposium
February 9, 2009

OVERVIEW

I. A Few Preliminaries
II. A Theory of Consciousness
III. A Theory of Mental Qualities
IV. Replies to Some Objections
I. A Few Preliminaries

- My concern this afternoon is *what it is for mental states to be conscious*—i.e., the *difference* between mental states that are conscious and those that are not.
- I’ll advance a theory of what it is for mental states to be conscious—But I’ll be concerned more specifically with what it is for *qualitative* mental states to be conscious—states such as sensations, perceptions, and emotions.

Explaining what it is for a mental state to be conscious requires explaining how the conscious cases differ from those that are not conscious. So it’s crucial that some mental states (of whatever type we’re concerned with) do occur *without being conscious.*

- It’s widely acknowledged that states such as thoughts, hopes, doubts, desires, and the like do occur nonconsciously.
- But many theorists still resist the idea that *qualitative states,* such as sensations and perceptions, can occur without being conscious.
But there is strong experimental evidence for *subliminal, nonconscious perception*. E.g., a visual stimulus presented in such a way that one is unaware of seeing it will often affect one’s mental processing.

These so-called *priming effects* can concern *visual qualities* such as color and shape. So these effects are presumably mediated by qualitative mental states—albeit states that aren’t conscious.

Also “blindsight” patients, with damage to primary visual cortex (Weiskrantz 1997), guess highly reliably about the visual qualities of stimuli that they’re *unaware of seeing*.

There is also evidence of nonconscious *change detection*—showing that there are post-change sensations of which subjects are wholly unaware (Fernandez-Duque and Thornton 2000a, b, 2002, Silverman and Mack 2006, Laloyaux, Destrebecqz, and Cleeremans 2006).

Basic both to experimental work and to common sense is *an intuitive idea about what it is for a state to be conscious*:

A mental state is conscious only if one is in some way aware of being in that state; otherwise it is not conscious.

I’ll call this the *Transitivity Principle* (TP).
As just stated, TP has the air of a trivial commonsense observation: Just as one smiles a smile, so one experiences an experience (Sosa 2003; Block forthcoming).

But this “internal accusative” formulation *in effect assumes that experiences are always conscious*, which gives us no help about the difference between states that are conscious and those that are not.

So this deflationary understanding can’t do justice to experimental findings that subjects are sometimes in qualitative states they aren’t in any way aware of—and hence don’t count as conscious states.

So a better way to put TP is this: If one is *not aware* of some mental state that one is in, that state is *not* conscious.

These two versions of TP are of course logically equivalent. But putting it this way *makes it clear that qualitative states need not be conscious*—and reflects experimental practice in looking for qualitative states that subjects are unaware of.

In §II, I’ll build on TP to formulate a theory of what it is for mental states to be conscious.
But first let me mention a distinction due to Ned Block between two ways one might think of mental states’ being conscious.


By contrast, a state exhibits **phenomenal consciousness** if there is something it’s like to be in that state.

But there’s reason to doubt that either notion **conforms to TP**. I’ll address access consciousness now, and phenomenal consciousness in §IV.

As Block has noted (2007, 2008), a state’s being access conscious means its content is in a global workspace located in frontal cortex—making it available “for use as a premise in reasoning, … [and] for [the] rational control of action and … speech.”

But content can be in the global workspace and so can be thus available **without one’s being in any way aware of that state**.

Signal strength ensures that one will **typically** be aware of any states that’s thus available. **But it’s the awareness that matters for the state’s being conscious, not its being in the global workspace.**
II. A Theory of Consciousness

- Explaining what it is for qualitative mental states to be conscious requires two things:
  1. An explanation of what it is for a state of any sort to be conscious; and
  2. An explanation of what it is for a state to be qualitative in the first place.

- I'll advance a theory of the first in this section, turning in §III to a theory of the second.

As I've argued, a satisfactory theory must conform to TP. And we've seen that not all theories do.

- Block's notion of access consciousness doesn't implement TP. Similarly for Fred Dretske's "first-order" theory, on which a state's being conscious consists in its making one conscious of something.

- These theories have difficulty with mental states that aren't conscious, since states can make one conscious of things without the states themselves being conscious (contra Dretske), and can also be poised for the effects Block lists.
TP says that no state is conscious unless one is in some way aware of that state. So the crucial question for any theory of consciousness is: 

\textit{In what way is one aware of a state when that state is conscious?}

We’re conscious of things in two ways: (1) By \textit{sensing or perceiving} those things, and (2) by \textit{having a thought} about those things (as being present to one). Which of these ways implements TP?

The most widespread theory (Aristotle, Locke, Kant) is (1)—\textit{inner sense}: We’re aware of conscious states \textit{by sensing them}.

But there is no such inner sense. We \textit{individuate} each sense by a distinctive range of mental qualities—color, sound, etc. And there are no distinctive qualities that could figure in some inner sense.

(Aristotle wrongly identified mental qualities with properties we perceive, and so held that inner sense and ordinary sensing involve the \textit{same} qualities.)

Since we don’t sense our conscious states, it must be that we’re aware of them by having thoughts about them. Since those thoughts are about other mental states, I’ll call them \textit{higher-order thoughts} (HOTs).

This process of elimination is not the only reason to posit HOTs; I’ll mention a few.
In both commonsense and experimental contexts, we determine whether a state is conscious by whether it’s reportable—by whether the individual can tell us about it. And all speech acts express thoughts with the same content. So saying that one is in a state expresses a thought that one is in that state—a HOT. And being able to report a state requires the HOT that the report would express. We can best explain why a state’s being reportable coincides with its being conscious by positing that conscious states are accompanied by HOTs.

We’re aware of our conscious states in a way that’s subjectively unmediated and subjectively “from the inside.” So our HOTs mustn’t rely on any inference or observation of which one is aware. A state is conscious if one is aware of it; so it’s not that the HOT transfers its own consciousness to the state it’s about. Indeed a mental state can make one aware of things even when it’s not a conscious state; subliminally perceiving something, e.g., makes one aware of it—though not consciously aware. So the HOT need not itself be conscious.
It’s therefore not a problem that we’re seldom aware of any HOTs. Indeed, the theory predicts this: We’ll be aware of a second-order thought only if there’s a third-order thought, which is itself rare.

Support for the theory comes not from introspection but, as with other theoretical posits, from the range of data—such as TP and the reportability test—that the theory explains. We’ll see more such explanations in §III.

There’s more to say about the nature of HOTs—some to emerge in §§III and IV. Graphically, then,
I have a thought that $p$
It’s sometimes argued that HOTs may explain the consciousness of some kinds of mental states, but not all. Perhaps, it’s argued, a thought’s being conscious consists in one’s having a HOT about it, but not so for sensations and other qualitative states. Sensations are conscious in respect of their distinctive qualitative character, and it may seem that HOTs can’t capture that.
One apparent difficulty is that HOTs have only conceptual content, whereas conscious sensations differ in qualitative character, which varies in ways that greatly outstrip our concepts. We have no words—hence no concepts—for all the mental qualities that conscious sensations can exhibit.

That’s a major reason why the inner-sense theory has long seemed so tempting.

But since there are no higher-order mental qualities, we have to see whether the HOT theory can meet this difficulty.

We describe our sensations by borrowing words used for the physical properties that our sensations enable us to perceive—e.g., colors and sounds.

We use ‘red’, ‘square’, ‘loud’, and ‘high pitched’ to describe both our sensations and the physical stimuli that cause them.

These words mean something different when applied to mental qualities and to physical stimuli: Visual sensations aren’t red in the same way that tomatoes are.

This borrowing of terms points to the way we actually describe—and conceptualize—the mental qualities of our sensations.
Perceiving things requires discriminating among perceptible properties—the properties an individual can perceive. And to do that, it must be that the individual is in states that resemble and differ among themselves in ways that correspond to the perceptible similarities and differences among the relevant objective physical properties.

The relevant similarities and differences among perceptible properties are those that are perceptible by the individual. Differences among mental qualities echo perceptible differences among stimuli.

Mental qualities are the differential properties of sensations that enable us to discriminate among the physical properties accessible by each perceptual modality. Since discrimination among perceptible properties sometimes occurs subliminally—and hence without being conscious—a state’s having some mental quality is independent of its being conscious.

So we can individuate mental qualities independently of their being conscious—by their position in a quality space that corresponds to the quality space of the relevant perceptible stimuli.
This CIE graph locates all chromaticities (saturation and hue)—hence all corresponding color sensations.

Monochromatic hues on the outside (in nm); less saturated hues inside the spectral locus.

(Commission Internationale d’Éclairage)

This device of identifying mental qualities by appeal to their location in a quality space that corresponds to the locations perceptible properties have in their quality space is not new (see, e.g., Wilfrid Sellars, Science, Perception and Reality, chs. 2, 4, 5, and Sydney Shoemaker, “Functionalism and Qualia”).

What is new is the use of this technique to give an account of mental qualities that’s independent of their being conscious (Rosenthal, Consciousness and Mind, chs. 5-7, esp. 7).

Because it’s independent of consciousness, this account fits well with an appeal to HOTs to explain the consciousness of qualitative states.
The HOTs that accompany qualitative states, and make one conscious of them, describe those states in terms borrowed from descriptions of corresponding stimuli.

Since location in a quality space doesn’t require a concept for each mental quality, we can solve the “outstripping” problem.

HOTs will use the usual repertoire of concepts for qualities—red, green, etc.—and can then describe small variations by relative location in the relevant quality space: closer to this color, darker than that. And that’s how we do describe fine-grained differences in what we perceive.

And there is evidence that this is how we are aware of fine-grained differences among our conscious qualitative states.

We’re far more likely to be conscious of two slightly different mental qualities as distinct when they occur together than one after the other (Halsey and Chapanis 1951).

The best explanation of this is that we’re conscious of fine-grained differences when the qualities occur together because we are conscious of by their comparative location in a quality space, and such comparative locating is a lot harder to do when the qualities don’t occur together.
There is a particular version of so-called change blindness that shows that qualitative states can be conscious in respect of mental qualities they don’t have.

James Grimes (1996) switched stimuli during saccades, when no retinal input reaches visual cortex. But even after saccades, subjects often didn’t consciously see the changed post-saccade stimuli, taking themselves still to be seeing only the pre-saccade stimuli.

But post-saccade stimuli do get through to visual cortex; so subjects’ qualitative states diverge from awareness of them.

There is a nice theoretical side effect to giving an account of mental qualities that’s independent of their being conscious.

If we could know about mental qualities only subjectively, from the inside, you could never know whether my mental quality of red is the same as yours—nor even whether I have any at all. These traditional problems of the inverted spectrum (Sextus Empiricus, Locke) and absent qualities would then be unavoidable.

We avoid those quandaries if, as on the quality-space theory, we know about mental qualities by way of perceptual function.
Summarizing so far: HOTs implement TP: A state is conscious if one is conscious of oneself as being in that state.

But we’re conscious of qualitative states in respect of their mental qualities; so we need an account of mental qualities to explain how we’re aware of qualitative states.

On the foregoing theory, we’re aware of mental qualities in respect of their location in a quality space, and HOTs represent qualitative states in that way.

And since mental quality is independent of consciousness on that theory, the combined theory conforms to TP.

IV. Replies to Some Objections

If being aware of a mental state is enough for that state to be conscious, why doesn’t that work generally? Why isn’t a stone conscious when one is aware of the stone (Dretske)?

This is a challenge not just to the HOT theory, but to TP itself.

This objection is better cast in terms of others of one’s states—e.g., states of the liver (Block 1995)—so as to sidestep the fact that objects such as stones cannot be conscious in the way mental states are.
States of the liver aren’t ever conscious; and it’s likely that a state’s being conscious requires that it have either *intentional content*, as thoughts do, or *qualitative character*, as do sensations.

And we don’t currently think of liver states as having such mental properties.

Still, if you were aware, *independently of any apparent inference or observation*, of your own liver states in qualitative terms—say, in respect of the bodily conditions those states represent—those states might then seem, subjectively, like a strikingly new kind of conscious state.

So it’s because we pretheoretically regard liver states as having no mental properties that we have the intuitive sense that being aware of such states in a suitable way would not result in their being conscious.

TP says that a state’s being conscious consists in *one’s being aware of oneself in a suitable way as being in that state*.

So being aware of somebody else as being in a mental state would not result in that state’s being conscious—even if one were thus aware in a way that’s subjectively noninferential and nonobservational.
Do nonlinguistic creatures (babies, dogs) have HOTs? If not, the theory has the unintuitive, unwanted consequence that their mental states are never conscious.

Language expresses one’s thoughts, but thoughts can occur without language. So nonlinguistic creatures may well have HOTs, though doubtless fewer and less complex than those of humans.

A more pressing challenge about human infants stems from the likelihood HOTs occur in prefrontal cortex (PFC), since it’s known that PFC is not especially well-developed in infancy.

But infants do have some frontal-lobe function (Csibra et al 2001, Diamond et al 1989, Bell 1998)—and so may well have some HOTs.

Also, though it might seem obvious from their behavior that infants’ mental states are often conscious, we can’t count on it.

Mental states can have the same effects on behavior—and other mental states—whether conscious or not. So behavior by itself doesn’t tell us whether infants’ mental states are conscious.

We can settle that for both nonlinguistic animals and infants only with a theory that deals well with clear cases (adult humans).
The intuitive pull to think that the mental states of infants and nonlinguistic animals are mostly conscious stems from the idea that if a state is conscious in our case it'll be conscious in theirs.

For many kinds of mental state, it's likely that to a situation that results in our being in that state also result in infants and nonlinguistic animals' being in it.

But we can't infer from our states' being conscious to theirs being conscious. We would extrapolate in that way only if we reject TP, thereby running together a state's occurring with its being conscious.

I've argued that TP is made true by distinct higher-order states. Following Aristotle, Caston (2002) has urged instead that each state is simply aware of itself.

But recall: We can more often discern a slight qualitative difference between two conscious states if they occur together than if they occur successively.

We can't explain this by appeal to any factor that's intrinsic to the states, since intrinsic factors won't vary depending on simultaneous or successive occurrence.

We need some extrinsic factor—such as distinct HOTs.
Block has argued that qualitative states exhibit a kind of consciousness *simply in virtue of their being qualitative*: A state has such phenomenal consciousness if there is something it’s like to be in it.

Block (1995) originally held that phenomenal consciousness occurs *even when one is in no way aware of the relevant state*—as in subliminal vision and various pathologies.

But recently (2007) he acknowledges that without some such awareness, *phenomenal consciousness would not be any kind of consciousness at all.*

This means endorsing TP. But Block now also holds we can have TP by appeal not to distinct HOTs, but to the *deflationary version* of TP that we always experience experiences and to Caston’s view that *states are reflexively aware of themselves*.

But neither of these moves works. As we saw in §1, Sosa’s deflationary idea that we always experience our experiences can’t draw the distinction between states that are conscious and those that are not.

And an “intrinsicalist” view that states are aware of themselves would posit explicit higher-order content no less than HOTs.
So it’s unclear what a state’s being phenomenally conscious amounts to apart from its just having qualitative character.

If this is indeed a kind of consciousness, then any state of the requisite kind—any qualitative state—will have phenomenal consciousness. So such consciousness cannot accommodate a distinction between a particular state’s being conscious and its not being conscious.

And lacking a contrast between being conscious and not, this is not the intuitive, commonsense kind of consciousness we’ve been concerned with.

I’ve argued that qualitative states sometimes aren’t conscious at all, relying on the reportability test that’s standard in experimental work and implicit in our commonsense folk judgments.

But Block argues that reportability goes with consciousness only if the one has suitable access to the state, and one can be aware of a state without having the access needed to report it.

Block urges that awareness occurs without access suitable for reportability by appeal to experimental results that stem from the pioneering work of George Sperling (1960).
Subjects very briefly shown a matrix of alphanumeric characters can identify only very few, but can identify a single row if cued right after about which row to report:

<table>
<thead>
<tr>
<th>2</th>
<th>B</th>
<th>3</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>U</td>
<td>0</td>
</tr>
<tr>
<td>Y</td>
<td>E</td>
<td>7</td>
<td>N</td>
</tr>
</tbody>
</table>

Subjects can identify the characters in any random row after cuing for that row.

Block infers that there is phenomenal consciousness of all the characters—and hence awareness all of the sensations of the characters—even given the limited ability to report.

But that’s not at all obvious. Subjects report seeing all the characters—they just can’t identify them all.

So we can explain the Sperling results by positing sensations that represent all the individual characters, along with awareness of those sensations that’s fine-grained enough to recognize all the items as alphanumeric characters—but not enough to identify them all.

If so, awareness of all the sensations goes with the access needed to report them—and awareness and reports vary together as regards how fine-grained they are.
So—apart from there being some specific, detectable impediment to linguistic performance—reportability of mental states goes with awareness of them.

Since we can rely on reportability, we can conclude that phenomenal consciousness consists after all only in mental qualities, independent of any awareness of those qualities.

Awareness of those qualities must occur by way of distinct states, such as HOTs.

Since phenomenal consciousness does not implement TP, it does not conform to our commonsense notion of consciousness.

HOTs have no qualitative character. They are responsible for qualitative states’ “lighting up” not by contributing their own mental qualities, but by making us aware of the qualities of states they are about.

There is independent reason to think that HOTs do affect conscious qualitative character. In learning to taste wines or to distinguish among musical instruments, it’s useful to have words to classify the new experiences. Having ‘oboe’ to label this mental quality and ‘clarinet’ for that can lead to the mental qualities’ being subjectively different. Also with wines.
How can learning words have that affect? It’s not credible that learning new words causes states with new mental qualities. What mechanism could explain that?

But words express thoughts. And since the words here describe one’s mental qualities, the thoughts are HOTs. So the learning of new words in these cases evidently make possible HOTs with new contents—and thereby make us conscious of our mental qualities in new, more fine-grained ways.

What HOTs we have evidently affects how our qualitative states are conscious.

Summary

TP and a HOT theory give us a defensible account of what it is for mental states to be conscious—an account that explains a number of central commonsense intuitions and experimental findings.

For the HOT theory to apply to qualitative consciousness, we also need a theory of mental qualities that’s independent of qualitative states’ being conscious.

The combined HOT and quality-space theories enable us to account for qualitative consciousness.
THANK YOU FOR YOUR ATTENTION