The nature of consciousness has become central in psychology, philosophy, and neuropsychology in recent years, with two annual interdisciplinary conferences, two interdisciplinary journals, and a raft of books all dedicated just to consciousness. Among questions of central interest are the neural correlate of consciousness—how neural events give rise to conscious states such as perceiving, thinking, and intending—and whether such mental states sometimes occur without being conscious. But it’s striking that much of this discussion and debate occurs without any theoretical account of what the target phenomenon—consciousness itself—consists in.
This would not occur in any other area of investigation. One can’t seriously debate issues about a phenomenon or investigate its scientific underpinnings without first having a moderately clear account of what that phenomenon consists in—even if the account is rough and subject to revision.

Many justify the lack of any such account here by claiming that consciousness is a special case: We all know well enough what it is in a subjective way. We’re all conscious, and all have conscious thoughts, perceptions, sensations, and feelings.

Why isn’t that good enough? Why doesn’t that specify the target phenomenon well enough to proceed?

It’s clear that it doesn’t—and can’t. Take one of the questions just raised: Do mental states occur without being conscious?

Our subjective inner lives not only can’t settle that; they have no bearing on it at all. We have access “from the inside” only to conscious states. So our subjective point of view can’t even address the question of whether mental states are all conscious.

Similarly for finding a neural correlate of consciousness: We can’t do that unless we have some independent, reliable fix on what consciousness is—what we’re looking for a neural correlate of. The absence of a theoretically informative account both impedes and biases research.
In my talk today I’ll develop an informative theoretical account of what it is for mental states to be conscious. I first consider a few commonsense considerations that any satisfactory theory must accommodate. I then build on those preliminary points to develop a theory of conscious mental states, along the way giving arguments in support of the theory and then some against its main theoretical competitors. I conclude by applying the theory to what many see as the most difficult test case: the way perceptions, bodily sensations, and other mental states exhibit conscious qualitative character. And I argue that, by that test, the theory does very well.

Overall Plan

I. The Phenomenon of Consciousness
II. A Theory of Consciousness
III. Qualitative Consciousness
1. The Phenomenon

- Nobody is more responsible for the recent interest in consciousness than Thomas Nagel, whose 1974 paper, “What Is It Like to Be a Bat?”, raised challenges that has provoked discussion ever since.
- Nagel urges that any objective account of consciousness would fail, since it would lose the distinctively subjective. But it’s crucial to highlight a distinction that Nagel’s discussion elides: between a conscious creature—a bat or a person—and a conscious state, e.g., a perception, thought, desire, and so forth.

- A creature is conscious if it’s awake and receptive to sensory input. But creatures are often conscious even if not all of their thoughts and perceptions are conscious. Subliminal perceiving isn’t conscious. And a great many thoughts and desires also occur without being conscious.
- And even if perceiving and thinking were all conscious, the two questions are distinct: What it is for a creature to be conscious and what it is for a state to be conscious.
- Another preliminary distinction: We speak of being conscious of things—as when we see or hear something or (sometimes) when we think about it. That’s the same as being aware of the thing.
And that’s again different from a mental state’s being conscious. If you see an object subliminally, you’re aware of the object—just not consciously aware of it.

One is conscious or aware of an object by being in a mental state that represents that thing. And subliminal perceptions represent stimuli, though they aren’t conscious. A state’s being conscious is not the same as one’s being conscious of something by being in that state.

Failing to distinguish these tacitly assumes that all mental states are conscious. Since all mental states represent things, if that by itself were enough for a state to be conscious, all mental states would be.

Armed with these basic distinctions, we can ask what it is for a state to be conscious—as against a creature’s being conscious or a creature’s being aware of something.

Suppose we have powerful evidence that a person sees or feels or wants or thinks something—but the person sincerely and firmly denies doing so. We might just take the person’s word for it and conclude that the evidence somehow misleads.

But the evidence might be overwhelming. So we might instead conclude that the person does indeed see or feel or think the thing, but that the relevant mental state is just not a conscious state. The person sees and so forth—just not consciously.
This is indeed a wholly natural conclusion. Novels and plays are filled with examples of this—of characters who want or feel or think something but would sincerely and firmly deny doing so.

Even ordinary life we sometimes see what somebody thinks or wants when the person is wholly unaware of doing so. And there is robust experimental evidence of subjects’ seeing or hearing things unconsciously.

This gives us a first step toward a positive account of what it is for a mental state to be conscious: Sincere denial reflects that one is unaware of being in a state. So a mental state one is unaware of being in is not conscious.

And that means a state is conscious only if one is in some suitable way aware of being in it. Being aware of a state is a necessary condition for that state to be conscious.

To get conditions that are both necessary and sufficient for a mental state to be conscious, then, all we need to do is figure out what that “suitable way” is in which we’re aware of being in a state when that state is a conscious state.

Having that would give us a theory of what it is for a mental state to be conscious. It would tell us under what conditions a mental state counts as being conscious. And that in turn would help resolve many pressing theoretical disputes.
This is all clear even before we start to theorize. A mental state of which one is wholly unaware is not conscious. But we also know, prior to theory, some aspects of how we are aware of conscious states.

Our awareness of conscious states seems subjectively to be independent of any self-observation or inference about oneself—and so independent of what others say about us or what theories tells us.

Let me stress: “seems subjectively.” One’s awareness of some conscious state might indeed rely on self-observation or inference of which one is wholly unaware. But we have no way to know whether it does going just on subjective impressions.

Traditional theorists, like Descartes, thus overstepped when they insisted that we are directly aware of our conscious states. We know only that our awareness does not seem mediated, not that it actually isn’t.

In addition, our awareness of our conscious states seems spontaneous and to occur without any voluntary input from us. This contributes to the sense of immediacy.

And when a mental state is conscious one is aware not just of the state, but of oneself as being in that state. Otherwise we would lack any sense that the mental state is a state of oneself. It would be like being aware of a disconnected mental state—not belonging to any individual.
II. The Theory

- It’s entirely natural to think about things in terms of prototypes. But doing so also often misleads, and led to a serious wrong turn in the traditional picture about what type of awareness figures in mental states’ being conscious.
- The paradigm of being aware of something is sensing or perceiving that thing—seeing or hearing or touching and so forth. And that led many (e.g., Locke, Kant, and a few contemporaries) to posit some inner sense in virtue of which we’re aware of our conscious mental states.

- Positing an inner sense is fine if that’s just a metaphor for whatever way we’re aware of our conscious states. But taken literally it can’t be right.
- Sensing involves qualitative character—properties like color, sound, and so forth. Even theorists who deny that there is any mental qualitative character allow that perceptible properties, such as physical colors and sounds, figure in all sensing.
- But no mental qualities and no perceptible properties figure in the way we’re aware of our conscious states. That’s especially clear with conscious thoughts. But even in conscious sensing we do not literally sense our sensations and perceptions.
And the commonsense ideas we surveyed earlier already pointed to the kind of awareness that is relevant. Recall our having overwhelming evidence that somebody is in some mental state, though the person sincerely denies that.

We concluded that the person is in that state, but that the state isn’t conscious. So a conscious state is, typically at least, a state an individual can report being in. Sincere report is a reliable indicator that the state being reported is conscious. We use that indicator in everyday life, and it’s the standard test in experimental work.

But why is that. What is it that explains why reportability is such a reliable mark?

Let’s assume that the reports in question are sincere. Reports are statements—and so reflect what the speaker thinks. If I say that it’s raining, e.g., I express my thought that it’s raining.

So if I say that I see or hear something or have some thought or desire, I also express my thought that I see or hear that thing or have the thought or desire I describe.

A state is conscious if one is aware—in a suitable way—of being in that state. And a sincere report is a reliable mark that the mental state reported is conscious. Since that report expresses a thought that one is in the state, having such a thought must be how one is aware of the state.
Let me run through that again. No mental state is conscious if the individual sincerely denies being in it. Typically, a state is conscious only if one is prepared to acknowledge being in that state.

Sincere denial shows that one is unaware of the state. So state is conscious only if one is in some way aware of being in it. But what kind of awareness of the state figures in the state’s being conscious?

Just as denial shows unawareness of the state, a report indicates awareness of it. But reporting state also expresses one’s thought that one is in the state. So one's having such thoughts best explains the way one is aware of one’s conscious states.

We can call the thoughts in virtue of which one is aware of one’s conscious states higher-order thoughts (HOTs). Being “higher order (HO)” here means just that the thought is about another mental state.

As noted before, we must not be aware of any self-observation or inference that leads to such HOTs, since otherwise our being aware of the state would not seem direct.

And the content of a HOT will be, “I am in such-and-such a state”—to ensure that one is aware of oneself as being in the state. HOTs will rarely themselves be conscious; indeed, we are seldom aware of them. A HOT would be conscious only if one had a third-order thought about that HOT.
Introspection is the special case in which we do have such third-order thoughts; then we are aware of our HOTs. But HOTs are in the first instance theoretical posits. We know about them as we know about electrons—as posits that explain the data, in this case, about conscious states.

And positing them constitutes a theory of what it is for a state to be conscious. A mental state is conscious if one has a HOT to the effect that one is in that state.

There are various additional fine points about the nature of these HOTs. But I’ll turn now to the main competitors of the HOT theory, and then consider a popular objection to the theory.

Fred Dretske has argued forcefully that any part of a conscious experience is itself conscious, but we are sometimes unaware of parts of conscious experiences. Such parts would be counterexamples to the claim that no mental state is conscious unless one is aware of it in some way.

He concludes that a state’s being conscious consists not in one’s being aware of it, but rather in its being a state in virtue of being in which one is aware of something. On his view, conscious states are states that one is conscious with, not conscious of.

But all mental states result in one’s being aware of things. So on that account no mental states would fail to be conscious.
Dretske has replied that no perception is conscious unless one can cite the fact one perceives. But citing a fact one perceives requires being aware of perceiving that fact. So we need HO awareness after all.

Dretske’s apparent counterexamples are, moreover, arguably cases in which one is unaware of a part of an experience in one way, though aware of it in another.

Many conscious states plainly figure in general psychological functioning. So some (e.g., Baars, Dehaene) have concluded that all do—that a state’s being conscious consists in its coming to be in a so-called global workspace, which makes it available to psychological functioning generally.

But there are counterexamples in both directions. Peripheral perceptions have negligible effect though they’re conscious. And many nonconscious beliefs and desires figure in general psychological functioning. Consciousness is not global availability.

Similarly, it’s not attention (various theorists). Conscious peripheral perception is rarely attended, and the experimental evidence of both spatial and object attention with no conscious awareness is now decisive.

Uriah Kriegel accepts that we need some HO awareness, but urges that it’s an aspect of the conscious state itself. But there is no reliable, independent way to individuate states that supports that claim.
An apparent advantage of Kriegel’s view is that it precludes a HO awareness from occurring on its own, with any target state. But Kriegel’s proposal doesn’t rule that out. A mental state could consist simply in an awareness that one is in a particular state, even though the mental state that consists in that awareness has no other mental properties—so that no target state occurs. Still, this points to a concern that underlies the most common objection to HO theories (e.g., Block, but also others): Can one be aware of oneself as being in a mental state that one is not in? Either because the HOT misrepresents the state or because the state just isn’t there at all?

Kriegel allows that the HO content might misrepresent the target state, but thinks the target must occur. But the difference is arbitrary: Whenever one misrepresents, what one represents isn’t literally there. HO theories allow for both, and both do actually occur. Dental patients sometimes seem subjectively to feel pain when there’s no relevant nerve. The best explanation is that they misrepresent their sensing fear and vibration as pain—so-called dental fear. When this is explained and drilling resumes, patients then report feeling vibration and fear—but still recall the earlier feeling as pain. Their memory is of the subjective awareness, not the target state!
Milder examples of misrepresentation by consciousness abound: Typically one consciously sees red things as generically red, though the first-order state typically exhibits a more fine-grained shade.

Such mild misrepresentation is harmless. More dramatic misrepresentation is rare because the first-order state is typically implicated causally in what HO awareness occurs—and because the system weeds out potentially harmful misrepresentation.

And there’s no problem even if the target state is wholly absent. A state is conscious if one is aware of oneself as being in it. Conscious states are how one’s mental life appears to one—need not be accurate.

III. The Qualitative

On the theory I’ve presented, each state is conscious in virtue of one’s having a HOT that one is in that state. Without such a HOT, the state would not be conscious.

But many hold that qualitative states—states such as sensations and perceptions, which have mental qualitative character—cannot fail to be conscious. Even Freud, who held that “[t]he mental, whatever its nature may be, is in itself unconscious” (S.F. XXII, 283), also insisted that “[i]t is surely of the essence of an emotion, that we should be aware of it” (S.F. XIV, 177).
If, as many hold today, being conscious is an aspect of mental qualities—intrinsic to qualitative states—then we must know about the nature of such states by the way they present themselves to consciousness.

So whatever knowledge we have about others’ qualitative states is secondary, readily overridden by their self-knowledge. This leads to various problems. There will be no way to detect whether your mental quality on seeing a red stimulus is the same as mine on seeing green (Locke)—nor even a way to say what that would amount to. And such a necessary tie with subjective access makes any connection with neural events seem problematic (Levine, Chalmers).

Most important, it’s then unclear what a qualitative state’s being conscious can—by itself—tell us about its mental quality. So if one sees being conscious as an aspect of mental qualities, their nature will seem ineffable. That’s why such theorists never offer any informative account of them.

These quandaries all result from taking the property of a state’s being conscious to be an intrinsic aspect of qualitative character. All qualitative states are then conscious. But the more damaging result is that the way mental qualities present themselves to consciousness is the first and last word about their nature. It overrules any other information we could have.
And all these difficulties simply vanish if we see being conscious not as an aspect of qualitative character, but as something added onto it—as the HOT theory holds.

We’d still learn about mental qualities from consciousness, but what we learn that way would no longer be the first and last word—only one kind of information among others.

And the most salient and useful information we ever have is the role that qualitative character plays in perceiving; the way qualitative states figure differentially in perceptual discrimination. That’s a highly rich and informative source of information. And it’s the only information first-person access actually provides.

Take discrimination of color, though the following considerations apply to all the perceptual modalities—and even to bodily sensations such as pains.

We consciously discriminate red from green stimuli, blue from yellow, and so forth. Doing so requires our being in states whose conscious mental properties differ in ways that correspond to the ways the stimuli differ for us. Only then would we be able consciously to discriminate the colors.

Here are the color discriminations the average person can make among the physical colors. This quality space maps the color properties we can discriminate.
But discriminating colors requires being in mental states that differ in ways that correspond to the differences among colors that we can discriminate.

So the very same color space also maps the mental qualities of color—those we use to discriminate physical colors. Relative location in that space gives us an account of what the mental qualities are—as other spaces do for the other modalities.

But perceptual discrimination isn’t always conscious, e.g., when it’s subliminal. So the quality space for the colors defines the nature of mental qualities both for conscious visual states and for subliminal visual states.

Conscious and nonconscious qualitative states have the same qualitative properties. They differ only in that the conscious qualitative states are accompanied by HOTs.

That tells us what the HOTs for qualitative states say: that one is in a state that has a particular relative location in the relevant quality space. So first-person access itself tells us only about perceptual-role. By allowing for qualitative states that aren’t conscious, the HOT theory accommodates an account of mental qualities that relies in no way on consciousness.

I’ll close by considering an objection based on the concern raised earlier about possible misrepresentation by consciousness (Levine).
Suppose I had a sensation of red but also a HOT that I have a sensation of green. What would that be like for me subjectively?

Subjectively it would be for me as though I had a sensation of green. Why?

A mental state is conscious if it seems subjectively that one is in that state—if one is subjectively aware of oneself as being in that state. (Recall dental fear.)

Consciousness is simply how my mental life appears to me subjectively. And that’s how I’m subjectively aware of myself as being in one mental state or another. Since my HOT determines how I’m aware of myself as being in various mental states, the HOT determines what it’s like for me.

Summary

HOT theory provides a powerful account of how conscious mental states differ from mental states that aren’t conscious, explains the pervasive use in experimental work and everyday life of reportability as a reliable mark of consciousness, and avoids the difficulties facing alternative theories.

The theory also accommodates qualitative states that aren’t conscious, and so makes room for an informative account of mental qualities, freeing us from the uninformative reliance on consciousness and first-person access to tell us about their nature.
Thank you for your attention!