

SUBJECTIVE CONSCIOUSNESS REDUCED?

Berit Brogaard

Uriah Kriegel. *Subjective Consciousness: A Self-Representational Theory*, Oxford: Oxford University Press 2009, 352 pages

What have we learned so far from studies of consciousness about the nature of consciousness? The answer, I fear, is nothing. We all have a great sense of what it feels like to be conscious yet none of us knows what it is. We don't even know if there is an answer to the question.

The majority of philosophers and scientists, wisely perhaps, leave the hard problem aside and tackle issues more likely to have answers. Neuroscientists are in the habit of addressing questions of what the neural correlates of consciousness are, and philosophers frequently engage in debates about whether conscious states supervene on, or perhaps even are identical to, physical states. Questions concerning the contents and language of consciousness have also gained popularity in recent years.

Only once in a while do we see a courageous attempt to solve the hard problem. Seen in this light, Uriah Kriegel's *Subjective Consciousness: A Self-Representational Theory* is a welcome addition to the long list of publications on consciousness. Besides being courageous, it is a book worth every penny. It is nicely organized, the language is succinct and the literature coverage is tremendously useful.

In line with tradition, Kriegel takes consciousness to be the "what it is like for me"-ness of conscious mental states, or what is sometimes called the 'phenomenal character' of conscious mental states.

Kriegel, however, divides phenomenal character into two components: qualitative character and subjective character. If I have an experience of something red, the phenomenal redness of my experience is the qualitative character of my experience, whereas the for-me-ness is the subjective character. Kriegel states that the qualitative character of conscious mental states determines which kind of conscious states they are, whereas the for-me-ness of conscious mental states is what makes them conscious in the first place.

According to Kriegel, what makes our inquiries into the nature of consciousness particularly thorny is that we don't always draw a line between qualitative character and subjective character. It's the for-me-ness that presents the main impediment for theories of consciousness, not the qualitative character.

The question of how brain processes can produce purely qualitative redness, Kriegel says, is no harder to answer than that of how physical matter comes to instantiate color.

Kriegel proceeds to offer an account of the subjective character of conscious mental states. In a nutshell, the state representing itself constitutes the subjective character of the state. The self-representational nature of conscious mental states is a kind of peripheral awareness. If I have an experience of something red, I am focally aware of redness but I am peripherally aware of the experience itself.

Of course, it would be easy to refute self-representational accounts of consciousness if there weren't any further constraints. There are. I will return to these constraints below. In what follows, I will offer an evaluation of three aspects of Kriegel's account.

The Mystery of Qualitative Character

The first aspect I will focus on is the hypothesis that qualitative character is no more puzzling than an explanation of why physical matter instantiates color. As Kriegel puts it,

the (genuine) puzzlement over the fact that a bunch of neurons vibrating inside the skull is associated with a yellowish qualitative character is no different from that surrounding the fact that a bunch of atoms lurching in the void is associated with a yellow color. It is no more surprising that neurons can underlie yellowishness than that atoms can underlie yellowness.

Whether Kriegel is right about this, however, depends on what we take colors to be. If colors are frequencies of incoming light or surface spectral reflectance properties, then it is not difficult to explain how things can have colors. But if color is understood in this way, then people with blindsight have mental states with color content. They can reliably detect wavelengths but without any reported awareness or phenomenal character. How this is possible is not very difficult to explain.

Explaining the nature of qualitative character, on the other hand, is a nearly impossible task. How do we explain how physical information reaching retina and then undergoing processing in the mid-brain (the LGN) and the occipital lobe (striate and extra-striate cortical areas) can result in purely qualitative redness? How can the double opponent processes in striate cortex and the further processing in V₁, V₂ and the V₄/V₈ color complex result in purely qualitative redness? That is a real brain teaser, and I believe it is this kind of

conundrum that David Chalmers would subsume under the “hard problem of consciousness.”

Needless to say, if objects instantiate qualitative colors, then the puzzle of how neurons vibrating in the head can produce qualitative colors is no different from the puzzle of how fundamental matter can instantiate qualitative colors. But if color primitivism is wrong, then the question of how vibrating neurons can produce qualitative color remains a genuine conundrum.

Degrees of Consciousness

I have a further worry about Kriegel’s theory of subjective consciousness. His account appears to be at odds with plausible gradability theories of consciousness.

When people go partially blind because of damages to the eye, the optical nerve or higher areas of the brain, nerve activity in the relevant areas is weakened. In people with neglect, for example, the neglect weakens or degrades the representation of the stimulus to the extent that it cannot be consciously perceived.¹

In masked priming studies, a masked prime precedes a target. The target is perceived differently when the prime is appropriately connected to the target. When the prime is masked, it is not consciously perceived. The reason for this is that the mask interferes with the visual stimulus. The stimulus is perceived but the neuron activity is too weak to produce conscious experience.

If weakening of nerve activity is a gradual process, which is extremely plausible, then there is no sharp boundary between conscious and unconscious states.

On a very plausible view of consciousness, consciousness requires feed-back from (dorsolateral) pre-frontal cortical areas to V1. Through several cycles of feed-forward and feedback of information between striate cortex and pre-frontal cortical areas, the nerve signals stabilize and consciousness miraculously appears. The lack of consciousness in cases of neglect, extinction, masked priming etc. can be explained by noting that the neuron signals never fully stabilize in these cases. They remain too weak for consciousness to arise.

The stabilization theory of consciousness is well supported empirically. But it is in no way clear how Kriegel would accommodate these views. For him, consciousness is an on-or-off feature. Either a state is self-representational in the right way or it is not.

1 Neglect is a disorder of spatial attention that results in failure to report on and orient to stimuli occurring in the visual field corresponding to the lesion. Mild forms of neglect are called ‘extinction.’

On the gradability theories, on the other hand, consciousness is not an on-or-off feature. There can be weak conscious states as well as fully conscious states, depending on the strength of the neuron activity. So, Kriegel's theory is squarely at odds with plausible neuroscientific theories.

I wonder how Kriegel would propose to accommodate Llinas and Ribary's 40 Hz theory, a theory consistent with the theories just outlined? On the 40 Hz theory, very vivid conscious experiences involve c. 40 Hz-oscillations in the relevant areas of the brain, whereas less vivid experiences and a lack of experiences involve slower oscillations. Slower oscillations occur when you are doing routine tasks, when you sit passively in front of the television screen, when you are lying in your bed ready to go to sleep, when you are actually sleeping and when you are in a coma.

If the 40 Hz theory is correct, and this theory has lots of evidential backing, then there is no sharp cut-off between conscious and unconscious states. Some conscious experiences are weak, some are moderately strong and some are strong. But if consciousness is self-representation, then either a state represents itself or it doesn't.

A Reductive Metaphysical Theory of Consciousness?

Kriegel's self-representational theory of subjective character is a metaphysical theory: It aims at identifying the essence of the for-me-ness of conscious states. Metaphysical theories of consciousness are quite rare. Even theories that may look like metaphysical theories in aiming at describing the features that all conscious states have in common are not in fact metaphysical theories.

The 40 Hertz theory, for example, is not a metaphysical theory. Creatures in distant parallel universes could be conscious even if the underlying realizer has no vibrating neurons. The same goes for all other neuroscientific theories of consciousness.

Kriegel's theory stands out from this group of theories because it aims at identifying what consciousness is. However, I do not think it succeeds.

Kriegel characterizes self-representation as a kind of peripheral awareness of the mental state in question. But now take my current red experience. Remove everything I am not focally aware of. Is that a metaphysically impossible experience? I don't see why it would be. Perhaps actual human beings cannot be aware without peripheral awareness. But it seems to me that there could be creatures without peripheral awareness.

There is even some empirical evidence supporting the claim that human beings can have experiences with no peripheral awareness. In 1993 psychologist Russ Hurlbert reported the case of Fran, a woman with borderline personality disorder. Her inner experience frequently contained multiple visual images, occurring simultaneously, all viewed straight ahead in a physically impossible overlaying. Her images often lasted for hours uninterrupted.

Fran had no figure-ground phenomenon in her inner images or in her outer perception. She experienced an entire visual scene without focusing on any of its aspects. No part of the image would be closer or in better focus than any other. She had no experience of zeroing in or losing focus. When Fran viewed the classical ambiguous duck-rabbit and face-vase figures, they would not alternate in her experience. She saw both aspects at once.

Fran had three television sets in her home and would watch all of them at once. Fran worked at a bank counting stacks of money, and she could count money and be deeply engaged in conversation at the same time. None of her co-workers could do that.

Fran's unhappiness was partially grounded in the problem that when therapists would try to teach her to think positively, she could do it, but she would simultaneously think negatively.

What should we say about Fran in regard to the claim that self-representation is a kind of peripheral awareness of the state in question? It is almost certain that Fran did not have peripheral awareness of her mental states. She only had focal awareness of the world and of her inner experiences. So, if Kriegel's theory is right, then the self-representation of Fran's mental states must have been a kind of focal awareness, not peripheral awareness.

Let it be granted, however, for argument's sake, that Kriegel has succeeded in showing that all conscious mental states are self-representational. Does it follow that Kriegel has offered us a metaphysical theory of consciousness?

No, because self-representation is hardly sufficient for consciousness.

My philosophical zombie could have an unconscious mental state with the property of representing itself. But that would not make the zombie's mental state a state with subjective character.

However, we don't need to go to the land of zombies to find counterexamples to the theory that any conscious state is conscious in virtue of representing itself. My (former) *unconscious* belief "this very belief is a belief," represents itself. But it is (or was) unconscious.

Now, Kriegel admits that representation is not sufficient for consciousness. For self-representation to suffice for consciousness, it must be non-derivative, specific and essential.

The first condition separates self-representational mental states from self-representing language. The second rules out generic self-representation, as in “All the sentences in this commentary are English sentences”. The third rules out self-representation that is not epistemically necessary. For example, “my mother’s sister’s only niece is making a mess” necessarily represents me but the representation is not epistemically necessary.

But these further constraints do not remove the underlying problem. My *unconscious* belief “this very belief is a belief” represents itself non-derivatively, specifically, and essentially. Just as there is no ideally conceptual scenario in which “I” does not represent the individual in the center, so there is no ideally conceptual scenario in which a belief with the content “this very belief is a belief” does not represent itself.

Kriegel uses the notion of a display sentence to illustrate essential self-representation. He invites us to consider the following example:

Suppose you peacefully drive on a bridge, when you suddenly see the words ‘under construction’ painted on the road. It is most likely that, when you read these words, a *complete* thought occurs to you, a thought with the full propositional content that you would express by saying ‘this bridge is under construction’.

The sentence that expresses the thought “this bridge is under construction” is a display sentence. The display sentence consists of the bridge itself and the symbols ‘under construction.’ Kriegel regards the bridge as representing itself in the envisaged example.

But this way of illustrating essential self-representation raises a further concern. When you form a mental image of your living room, you rely on memory. The image was stored in memory and was unconscious until you consciously decided to recall it. Moreover, it was stored with the words ‘an image of my living room’ painted on it. The stored image and the words painted on it is a display sentence. The stored image represents itself as being an image of your living room. So, the image is non-derivatively, specifically and essentially self-representing but it wasn’t conscious while it was stored. Kriegel’s theory predicts otherwise.

Here is another example of the same phenomenon. Since I was a child I have had vivid visual images in response to fearful or uncomfortable thoughts or pure discomfort. The images take the form of highly wrinkled bluish-greenish paper with sharp edges moving around in an irregular pattern. Sometimes they consist in rapidly presented irregular and wrinkled pieces of bluish-greenish cloth swirling around very quickly. Not all of my uncomfortable or fearful thoughts are accompanied by this sort of phenomenology but the occur-

rence of this kind of phenomenology is a sure sign of uncomfortable or scary thoughts.

For me, this bluish-greenish phenomenology *represents* fear, just as the word “cat” represents cats, or the word “Brit” represents me. For me, the bluish-greenish phenomenology together with my uncomfortable thoughts form a kind of display sentence, just like the bridge and the words ‘under construction’ do.

Sometimes I have an experience with the unmistakable bluish-greenish phenomenology without any direct awareness of fear. But over the years I have learned that when the bluish-greenish paper appears in my mind the fear is still there in the form of unconscious thoughts or emotions.

In this case, the display sentence in my mind is composed of a “subject term”, which consists of an (unconscious) uncomfortable thought that represents itself and the bluish-greenish phenomenology that represents fear. However, while the uncomfortable thought, in the envisaged case, is non-derivatively, specifically and essentially self-representing, it is not conscious. Kriegel’s theory predicts otherwise.

These considerations seem to show that Kriegel hasn’t given us quite what he promised. He has identified a law-like connection (which admits of exceptions) between self-representation and conscious states realized in the human brain, not a metaphysical theory of consciousness. But that by itself is quite impressive.²

2 I am grateful to David Chalmers for comments on an earlier version of this review. The review is based on my contribution to a book symposium on Uriah Kriegel’s book forthcoming in *Philosophical Studies*.