

Comments on Uriah Kriegel's /Subjective Consciousness: A Self-Representational Theory/,
Author Meets Critics Session, Pacific Division Meeting of the APA, 2010.

ARE CONSCIOUS STATES CONSCIOUS IN VIRTUE OF REPRESENTING
THEMSELVES? On Uriah Kriegel's *Subjective Consciousness: A Self-Representational Theory*

Berit Brogaard
February 24, 2010

This is an excellent book, one of the best I have read on consciousness in recent years. It is rigorously argued and contains interesting suggestions as to how to solve the mystery of consciousness. Following the standard literature, Uriah Kriegel takes consciousness to be the "what it is like for me"-ness of conscious mental states. This is also what is sometimes called the 'phenomenal character' of conscious mental states. For Kriegel (as for Levine), phenomenal character has two components: qualitative character and subjective character. If I have a red experience, the phenomenal redness of my experience is the qualitative character of my experience, whereas the for-me-ness is the subjective character of the experience. Kriegel states that the qualitative character of conscious mental states is what makes the conscious mental state the kind of conscious experience it is, whereas the for-me-ness of the mental state is what makes the mental state a conscious state in the first place. The mystery of consciousness, he says, does not lie in the qualitative character of experience. According to him, the question of how brain processes can give rise to purely qualitative redness is no harder to answer than that of how physical matter can instantiate colors. The mystery of consciousness lies in the for-me-ness of conscious experience. Kriegel then goes on to give an account of the subjective character of conscious mental states. He argues that the subjective character of conscious mental states consists in the state representing itself. It's the self-representational nature of conscious mental states that makes them conscious. The self-representational nature of conscious mental states is a kind of peripheral awareness. If I have a red experience, I am focally aware of redness but I am peripherally aware of the experience itself. Along the way Kriegel rebuts a number of alternative theories of subjective character: among others, that it is a kind of primitive property of conscious mental states, and that it is a representation of the conscious mental state by a higher-order state. I agree with many of Kriegel's arguments against both the naive primitivist view and the higher-order theoretical approach. My main concerns lie elsewhere. My three main points of disagreement can be summarized as follows: (1) Assuming that it makes sense to separate qualitative and subjective character I believe that the qualitative character of conscious mental states is at least as mysterious as the subjective character. (2) I believe Uriah's theory is at odds with plausible gradability theories of perception. (3) I am skeptical about the project of developing a reductive metaphysical theory of consciousness in terms of self-representation. I will consider these issues in turn:

(1) Granting for the moment that subjective character is a mystery, I think there is another phenomenon that is equally mysterious. According to Kriegel, "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". But that is exactly it. It is not surprising that light can have a certain frequency but it is surprising that purely qualitative redness can arise from physical information reaching our retina and then undergoing information processing in the mid-brain (the LGN) and the occipital lobe (striate and extra-striate cortical areas). How can the double opponent processes in striate cortex and the further processing of the information from V1 and V2 in the V4/V8 color complex possibly give rise to purely qualitative redness? That is a real mystery, and I believe this just /is/ an instance of the hard problem, in Chalmers' sense. Of course, if the world itself instantiates qualitative colors, which I happen to believe it does, then I agree with Kriegel. The puzzle of how neurons vibrating inside the head can be associated with a qualitative redness is then no different from the puzzle of how fundamental matter can be associated with qualitative redness. But if color primitivism is wrong, then the mystery of how qualitative color can arise from vibrating neurons remains a genuine mystery.

(2) One problem for Kriegel's actual theory of consciousness is that it is hard to see how he can account for the, in my opinion, very plausible gradability theories of perception. On one scientific theory of vision, what happens when people go partially blind is that the 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* [* 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']. 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, the prime 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 give rise to conscious experience. On a variation on this view, 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 primes etc. can be explained by noting that the neuron signals never fully stabilize. They remain too weak for consciousness to arise. These scientific theories of the correlates of conscious are exceedingly plausible, but Kriegel's theory seems to be conflict with them. For him, consciousness is an on-or-off feature. Either a state is self-representational in the right way or it is not. 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 forces us to reject plausible neuroscientific theories.

Kriegel's theory is also in conflict with Llinas and Ribary's 40 Hz theory (another gradability theory which is consistent with the theories just described). On this theory, very vivid conscious experiences involve ca. 40 Hz-oscillations in the relevant areas of the brain, whereas less vivid experiences involve slower oscillations. Delta waves are oscillations in the 0-4 Hz range, Theta waves are oscillations in the 4-7 Hz range, Alpha waves are oscillations in the 7-12 Hz range, Beta waves are oscillations in the 12-35 Hz range, and finally Gamma waves are oscillations in the 35-80 Hz range. Unconscious states (e.g. dreamless sleep) and very weak conscious states (e.g. deep levels of hypnosis or meditation and out of body experiences and other paranormal experiences) vibrate in the Delta/Theta range. Conscious experiences had during meditative states, under hypnosis and while relaxing, daydreaming or mind-wandering or while watching something boring on television are in the Alpha and Beta range, and intensely focused thought and vivid experiences (including dreams and emotional experiences) are in the Gamma range. Gamma waves can also be observed in cases of anxiety and hyperactivity. Francis Crick once called 40 Hz oscillations the "neural correlate of consciousness". But this claim is at best misleading and at worst false. 40 Hz oscillations are at best a neural correlate of extremely vivid experiences and intense thoughts and emotions. The 40 Hz theory has received a lot of empirical support in recent studies, especially after researchers started controlling for the oscillations of muscles and saccadic eye moments, which were thought to interfere with the measurements in the original studies. Recent studies by biophysicist Beverly Rubik and colleagues have shown that one can consciously alter the amount of one's 40 Hz brain oscillations compared to baseline through neuro-feedback. In neuro-feedback the EEG output from the prefrontal cortex is measured and the subject is then told or shown what the output is. The subject then consciously seeks to increase neuron oscillations by concentrating on a particular topic (for example unconditional love). The subjects in Rubik's study were meditators and non-meditators. The meditators were able to increase their 40 Hz oscillations 23-36%, whereas the non-meditators were able to increase their 40 Hz oscillations 10%. However, Rubik didn't find any lasting difference between the meditators and the non-meditators. The subjects in Rubik's studies all used mystic, love or happiness metaphors to describe their brain states at 40 Hz, for example, "detachment from myself", "trust in something bigger", "letting something higher come into my consciousness", "joy and bliss", "being in the moment", "thinking of people I love", "connection with the angelic realm", "pure love", "child-like wonder", "a loving feeling", "feels close to center of brain", "clairvoyance". When the subjects were asked to rate how well various adjectives applied to their Gamma wave states, "happiness" and "bliss" were given the highest rating. "Love" was given the second-highest rating. Through neuro-feedback the subjects were increasing their happiness and love.

Rubik also showed that psychedelic drugs can create abrupt changes in the 40 Hz brain oscillations. One 28 year old female subject volunteered to smoke the drug Salvia Divinorum, a psychedelic herb used in shamanic ceremonies by indigenous healers in Mexico. The subject had significantly less 40 Hz oscillations in the prefrontal cortex compared to baseline (her normal state) just as she reported peak intensity in her inner psychedelic state and significantly more 40 Hz oscillations compared to baseline just after she reported peak intensity. Salvia Divinorum creates distortions in body space perception. The subject in Rubik's study experienced herself as a crayon!

Emotionally hyper-sensitive people (including people with narcissistic personality disorder, OCD, asperger's syndrome and borderline personality disorder) also tend to have more 40 Hz brain oscillations than emotionally stable and emotionally insensitive people (e.g. schizophrenics). The amount of 40 Hz brain oscillations can be actively decreased with therapy or serotonin-reuptake inhibitors.

If the 40-Hertz theory is correct, and the theory is exceedingly plausible, then there is no sharp cut-off between conscious and unconscious experiences. Some conscious experiences are weak, some are strong. But if consciousness is self-representation, then there must be a sharp cut-off.

(3) Kriegel's self-representational theory of subjective character is meant to be a metaphysical theory. Metaphysical theories aim at providing the essence of the phenomenon under investigation. For example, a metaphysical theory of composition will provide an answer to the question of what it takes for several objects to compose a new object. Metaphysical theories of consciousness are not simply stating what is necessary given the physical laws of the universe or what is necessary given the metaphysical laws of the universe. For example, it is not part of the essence of time that it is actually raining today, even though it is metaphysically necessary that time is such that it is actually raining today. To give a metaphysical theory of a phenomenon is to tell us what it is to be that very thing. Now, as we are understanding 'consciousness' here, it is meant to refer to the what-it's-like-for-me-ness of conscious mental states. There are lots of theories of consciousness, so understood. The 40 Hertz theory is but one such theory. But the 40 Hertz theory is not a metaphysical theory of consciousness. Creatures in distant parallel universes with brains very different from ours could be conscious even if the underlying realizer has no vibrating neurons. The same goes for any other neuroscientific theory of consciousness.

Kriegel's theory differs from neuroscientific theories because it aims at saying what consciousness is. It aims to be a metaphysical theory. However, I do not think it succeeds as a metaphysical theory. Kriegel takes the phenomenal character of conscious mental states to consist in the conjunction of the subjective character and the qualitative character of conscious mental states. The self-representation of an experience is a kind of peripheral awareness. 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 human beings

cannot be aware of a red tomato without their experiences representing themselves. But it seems to me that there could be creatures who are aware of things in the world but whose experiences do not represent themselves.

There is empirical evidence that some people have experiences with no peripheral awareness. In 1993 psychologist Russ Hurlbert reported the case of Fran, a woman with borderline personality disorder, whose 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 Kriegel's self-representationalism? It is almost certain that Fran did not have peripheral awareness of her experiences. She only had focal awareness of the world and of her inner images. So, if Kriegel's theory is right, then the self-representation of Fran's visual images and experiences must have been a kind of focal awareness. But numerous data about Fran have been gathered through descriptive experience sampling, a method developed by Hurlbert, and the data do not reveal any focal self-representation. That, of course, does not mean there wasn't any self-representation. But it does indicate it. So, we might just have before us a strange case of a person who has /conscious/ images and experiences which do not represent themselves.

But let us even assume that Kriegel has succeeded in showing that all conscious mental states are self-representational. I am willing to grant for argument's sake that something like this could be right. Does it follow that Kriegel has offered us a metaphysical theory of consciousness? No, for self-representation is hardly sufficient for consciousness. I grant that phenomenological properties can be representational. My red experience has the qualitative-phenomenal property of representing something as red. According to Kriegel, it also has the subjective-phenomenal property of representing itself. These claims are interesting statements about the phenomenal character of experience. They tell us that the phenomenal character of experience is representational in two different ways. That is an interesting insight. But I do not see how telling us that my experience has these representational properties somehow tells us what it is like for me to have my current red experience. A zombie could have an unconscious mental state with the property of representing something as red and the property of representing itself as representing something as red. But that would not make the zombie's mental state a state with qualitative or subjective character.

But 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 being self-representational. Consider my (former) unconscious belief "this very belief is a belief". This belief represents itself. But it is (or was) unconscious. Now, Kriegel does recognize that self-representation is not sufficient for consciousness. The self-representation, he says, must be non-derivative, specific and essential. The first condition separates self-representational mental states from self-representing language. The second condition rules out generic self-representation, as in "All the sentences in this commentary are English sentences". The third condition 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. However, 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. Kriegel offers the following example of a display sentence: "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 this thought is a display sentence. Kriegel takes the bridge as representing itself in this sentence. He adds that all and only non-derivatively, specifically and essentially self-representing states are Mentalese display sentences (assuming that the Mentalese theory is right -- otherwise other constituents can play the same role). But now, this assumption is problematic. Consider the process of forming a mental image of your living room. When forming 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 form the image. 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. Hence, Kriegel's theory predicts that the stored image is conscious. But it isn't conscious until you recall it.

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 moving around in an irregular pattern. Sometimes the images consist in large quantities of quickly presented irregular and wrinkled pieces of bluish-greenish cloth moving around very quickly. Not all of my uncomfortable or fearful thoughts are associated with this sort of phenomenology but the occurrence of this kind of

phenomenology is a sure sign of uncomfortable or scary thoughts. When I was a child I used to be deadly scared of the moving wrinkled paper and cloth in my head. For me, this 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 I have learned over the years 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. But Kriegel's theory predicts otherwise.

In light of these considerations I think it is fair to say that at best Kriegel has succeeded in giving us 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 an achievement.*

* I am grateful to David Chalmers for comments on an earlier version of this commentary.