On the "feel" of things

The sensorimotor theory of consciousness

An interview with
Kevin O'Regan

By Cordelia Erickson-Davis

Kevin O'Regan
jkevin.oregan@gmail.com
Laboratoire Psychologie de la Perception
Université Paris Descartes, France

Cordelia Erickson-Davis
cred22@stanford.edu
Departments of Anthropology and Medicine
Stanford University, USA

You are primarily interested in the "what it's like" aspect of sensory experience. To address it, you've developed what is called the "sensorimotor theory" of consciousness (O'Regan & Noë 2001, Noë & O'Regan 2002; Myin & O'Regan 2002, O'Regan 2011, O'Regan 2014), which holds perception to be a law-governed mode of encounter with the environment. These laws are abstracted from the sensorimotor contingencies of the animal in relation with the environment—both the contingencies fixed by the perceiver's visual apparatus, as well as those fixed by the character of objects. These processes of sensorimotor interaction are distinct from perceptual consciousness, however, which you divide into "perceptual awareness" on one hand (or "transitive perceptual consciousness"), which is the exercise of one's practical knowledge of these sensorimotor contingencies, and "general perceptual consciousness" on the other, which is the capacity to become aware. Do I have that right? Could you further describe the theory, and in particular focus on how it addresses consciousness from the perspective of what you and others have described as the "easy" and "hard" problems of consciousness?

Yes, I'd say you've got it more or less right. But the way you say it sounds very technical. I would have liked you to stress the luminous simplicity of the idea and why it is a breakthrough in understanding consciousness!

To explain better, let me first note that most people think that consciousness is a mystery—they think that there is a problem in explaining how physical and chemical processes in a brain could somehow generate subjective experience. Philosophers call this the "hard" problem of consciousness.

Science doesn't just advance by making discoveries: it advances by defining terms in more precise ways. I think such redefining is what's needed to understand consciousness. The following is a way of defining consciousness that captures what most people mean by the term, and that at the same time dissolves the mysteries.

The definition has two layers:

First, at the top layer: in the normal everyday sense of the word, when you say you are conscious of something... there has to be a "you" with sufficient cognitive capacities. People would not normally say of a fly that it is conscious of the cheese it landed upon: The fly is presumably just a biological machine that is reacting to the environment. What about the mouse that ate the cheese? And the cat that ate the mouse? And the dog that chased the cat? And the child that chased the dog? And the adult that scolded the child? Clearly as we go higher in the hierarchy, we have higher degrees of being "conscious of". Certainly the adult's, if not the child's, understanding of the situation involves them not just reacting, but a variety of other things like knowing that they are reacting, knowing who "they" are, and knowing why they are reacting, and knowing that they know that they are reacting...

Though maybe too complex for flies and mice, there is nothing magical about such self-referring cognitive states. Being "conscious of" something in this way is what the philosophers call the "easy" problem of consciousness. It requires a variety of highly developed cognitive capacities, including the ability to conceive of one's "self"—but this mode of being conscious of something is not a mystery. It is coming to my smartphone in the next decades.

That was the top layer. But now there is the bottom layer of consciousness. I can be conscious of *an experience*. For example, I can be conscious of the hurt of the pain, or the redness of a red sunset. They *feel like something* to me. It's not just that I'm thinking of them or aware of them, like I can think of a pain or of a red sunset. I actually *feel* them. What is it like to feel things, rather than not feel them like when you are just thinking about them? This is what the philosophers call the "hard" problem, or the problem of "qualia". They think it's hard because they see no way brains could generate feels.

> " We should conceive of the feel of a pain, and
> the feel of red, and all perceptual experiences
> as *ways of interacting with the world*. "

And that is where I think a redefinition helps. In fact, the redefinition I propose is perfectly obvious to the man in the street, who would never imagine that a feel could be generated by the brain. What after all, is the feel of driving a Porsche? Well, it's the way it *handles* when you swing around the corner, it's how it *reacts* when you press the accelerator and it speeds forward... it's *how you interact with it*. Similarly, we should conceive of the feel of a pain, and the feel of red, and all perceptual experiences as *ways of interacting with the world*.

At first counterintuitive for the scientist looking for brain mechanisms, this way of thinking about experience provides an exquisitely simple account of the "hard" problem: when you have an experience, the what-it's-like of the experience is constituted by *how you interact with the world* when you're having the experience. But this experienced quality is not *consciously* experienced unless you as a person are attending to it, making use of it in your rational thoughts, decisions, planning etc., in the way of being "conscious of" that I described in the top layer of my account.

In summary: you are having a conscious experience of red when, at the top layer, you are "conscious of" the fact that, at the bottom layer: you are currently engaged in interacting with the world in a way that is constitutive of the laws of redness. The top layer provides the awareness, the bottom layer provides the experienced quality.

What motivated you to think about these issues, and what was your training up until that point?

Ever since I was a child I wanted to make a machine that thinks. I started off studying physics, because I thought the brain could be understood using the methods of statistical physics that try to model the behavior of large numbers of interacting bodies. I then moved into experimental psychology, where I worked on eye movements and visual perception. I realized that there was a logical flaw in the way people think about perception: people assume that perception involves the brain making an internal representation of the world. But then: who or what perceives that internal representation? It was this realization that led me to postulate a "sensorimotor" theory, where perceiving involves interacting with the world, not making an internal representation.

Many aspects of your theory resonate with other approaches that fall under the umbrella of "embodied cognition". This includes the "ecological perspective" of Gibson (1966, 1979), the programs of "active" and "animate" perception (Aloimonos et al. 1988; Ballard et al. 1997), embodied artificial intelligence (Brooks 1991), autonomous systems (Varela & Bourgine 1992), and enactive perception and cognition (Thompson & Varela 2001; Varela et al. 1991). How does the sensorimotor theory compare with these other theories?

The sensorimotor theory is trying to address the "hard" problem of consciousness: Why do things feel the way they do? Why does "red" look "red" rather than "green". Why does "red" not sound like a bell? Gibson's ecological approach and embodied artificial intelligence are not trying to solve that issue, and are instead looking at the role of action in perception. To me this is not very exciting: action obviously improves perception because it provides more information to be gathered. But these approaches miss another important role for action, namely what it brings to an understanding of the experienced quality of sensory experience and consciousness.

Autonomous systems and enactive approaches are, on the other hand, addressing the issue of consciousness. They invoke action as an essential element in consciousness. However, my impression is that they think that there is something magical about action. They think that interaction with the world somehow instills consciousness into biological systems. Their appeal to interaction, and concepts like autopoeisis, seems to be an attempt to use mysterious notions to elucidate what they think is even more mysterious, namely consciousness. What these approaches seem not to have realized is that if we understand the what-it's-like of perceptual experience as being constituted by what we do when we interact with the world, then there is actually no mystery. In other words, autonomous systems and enactive approaches correctly invoke action, but they don't realize why it is that action solves the mystery of qualia. They seem to want to keep a mystery where there is no mystery.

> " Autonomous systems and enactive approaches seem not to have realized that if we understand the what-it's-like of perceptual experience as being constituted by what we do when we interact with the world, then there is actually no mystery. "

In your theory, phenomenological inquiry takes on an entirely tractable tone. As you say, "the subject matter of phenomenological reflection is not an ephemeral, ineffable, sensation-like momentary occurrence in the mind, but, rather, the real-world, temporally extended activity of exploring the environment and the structure of sensorimotor contingencies" (O'Regan & Noë 2001, p. 962). What kinds of phenomenological reflection have you utilized in your experimental work? That is, how have you operationalized subjective inquiry in your empirical approach?

For example, I'm very proud of the work we did with David Philipona on color. Color a priori doesn't seem to involve interacting with the environment. But by taking a sensorimotor approach to color, we were forced to postulate that the experience of color is necessarily rooted in what happens when you move colored surfaces around in different lights. This gave a completely new idea about what color perception is, and made interesting predictions about what it means to be a "pure" color. We found a surprising link to anthropological data about color naming, where we accurately predicted which color names should occur most frequently. A simple philosophical idea, the sensorimotor approach, provided a surprising scientific prediction.

As you discuss in your book, the consequence of assuming that experience derives from the rules that govern action-related changes in sensory input is that the "feel"

of perceptual modalities like visual experience should be obtainable via channels other than vision ("provided that the brain extracts the same invariants from structure") (O'Regan & Noë 2001, p. 956). That is to say: sensory substitution. Can you describe the sensory substitution work you've done over the years, and what you've learned from it?

Indeed, the sensorimotor theory predicts you should be able to see with your ears, for example, or hear with your skin, provided you use some technical tricks to recreate the same sensorimotor laws via alternate sensory channels.

With my collaborators over the last years we have looked at how visual information can be conveyed through auditory input, how auditory information can be conveyed through the skin, and how it might be possible to obtain an augmented "sixth" sense of magnetic North via hearing.

Our efforts have been somewhat disappointing. We have discovered that it is technically not so easy to provide the brain with the right sensorimotor laws. Furthermore, we have found that the adult brain seems to be less flexible than we had thought. Our latest efforts to make a tactile aid that helps hearing-impaired people better understand speech has proven much more difficult to realize than we had anticipated. It seems that adult humans have a hard time making the arbitrary links that we require between speech sounds and tactile patterns. It may be that we have not been doing things right however. Perhaps the problem is that up until now, we have not included a proper "action" component in our approach. Sensorimotor theory would suggest that this would be necessary.

> " We will have machines that interact with us in ways that gradually involve higher and higher levels of cognition, including meta-cognition. We will not hesitate to say that they are conscious. "

You have said you believe that it is possible to build a robot that "feels". Does that mean that you believe artificial intelligence and robots are now or will be considered conscious?

Definitely. If consciousness is just a word that describes certain capacities we have to interact with the world, then machines are already on their way to being conscious. As I said earlier, whether a mouse, cat, dog, child or adult is conscious is a matter of degree. Similarly, in the next few decades, we will have machines that interact with us in ways that gradually involve higher and higher levels of cognition, including meta-cognition. When such machines interact with us socially every day,

when they have levels of knowledge and (meta-)cognition approaching (or superseding!) ours, we will not hesitate to say that they are conscious. Furthermore, when machines interact with the world with their senses, they will have experiences just like we have experiences when we interact with the world. The experiences will be different of course, precisely to the extent that their modes of interaction are different from ours. But that is true of mice, cats, dogs, and children too.

You have also said, in conversation, that you think that our societal focus on consciousness as a rubric for ethical decision-making is a mistake. Can you say more about this? What kind of ethics of artificial intelligence and other forms of cognitive enhancement technologies (e.g., brain machine interface devices) do you think is needed?

If consciousness is not an all-or-none thing, and is just a matter of having certain capacities (and meta-capacities) to interact with the world, then consciousness is useless as a criterion for ethical decisions. And even if there were some objective criterion of consciousness, it would be a pretense to invoke it: civilizations have often denied ethical respect to various perfectly conscious groups. Slaves, women, certain ethnic and religious groups, have all been denied human rights at various times through history. Ethics is ultimately a matter of social agreement, and human societies must take full responsibility for the decisions they take about whom to give ethical rights to. Appealing to science is just hypocrisy.

How can sensorimotor theory be applied to "alternative" states if consciousness— e.g., dreams and hallucinations? On these perceptual states, you've written that you believe them to correspond to implicit knowledge and implicit expectation, based on prior perceptual experience. Do you think the study of these states has anything to contribute to consciousness studies, from a sensorimotor perspective?

I personally haven't worked on implications of sensorimotor theory for altered states of consciousness, as produced for example by drugs, trances or meditation.

Note that some critics of sensorimotor theory have claimed that the theory cannot account for dreams and sensory hallucinations, since these occur without any interaction with the world. But this is to misunderstand the theory. The theory says that the quality of an experience resides in what you do when you are interacting with the world. If, through drugs or dreams you are in the same state that you usually are in when you are interacting with the world in a "red" way, then you will experience red, even if now you happen not to be interacting with the world at all.

The critics sometimes go on to say: well, doesn't that show that the brain does generate experience after all, since you can get the experience without interacting with the world? My answer is that the brain *enables* the experience, since a brain is

necessary in order to interact with the world. But that doesn't mean that the brain *generates* the experience in any meaningful way: Experiences are not the kinds of things that can be generated. Experiences are modes of interaction with the world. This new way of thinking about experiences is hard for some people to embrace. But note that a similar change in point of view happened as regards the notion of life. It used to be thought that life was *generated* by a vital spirit. Modern biology redefined our notion of life. It now considers that life is not the kind of thing that is generated. Life is *enabled* by various physical and chemical mechanisms like respiration, reproduction, etc. It would be meaningless to say that any one or other such mechanism "generates" life. Life is a capacity that certain systems have to interact with the world. Experience is the same.

What are you spending most of your time thinking about these days, and what's next for you?

Modestly, since I think that consciousness is no longer a mystery, I'm trying to solve a problem that I think is currently a mystery, namely why humans seem to be able to understand things. Today's machine learning architectures can do good pattern classification if they are given mounds of examples, but they don't understand what they're doing. Humans seem to understand things... I'm trying to understand what it is to understand.

References

Aloimonos, J., Weiss, I., Bandyopadhyay, A. (1988). Active vision. *International Journal of Computer Vision.* 1(4), 333-356.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723-742.

Braund, M. J. (2008). The structures of perception: An ecological perspective. *Kritike: An Online Journal of Philosophy*, 2(1), 123-144.

Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, 47(1-3), 139-159.

Gibson, J. J. (1966). The senses considered as perceptual systems.

Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition.* Psychology Press.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5), 939-973.

Myin, E., & O'Regan, J. K. (2002). Perceptual consciousness, access to modality and skill theories. A way to naturalize phenomenology? *Journal of Consciousness Studies*, 9(1), 27-46.

Noë, A., & O'Regan, J. K. (2002). On the brain-basis of visual consciousness: A sensorimotor account. *Vision and mind: Selected readings in the philosophy of perception*, 567-598.

O'Regan, J. K. (2011). *Why red doesn't sound like a bell: Understanding the feel of consciousness.* Oxford University Press.

O'Regan, J. K. (2014). The explanatory status of the sensorimotor approach to phenomenal consciousness, and its appeal to cognition. *Contemporary Sensorimotor Theory*, 15, 23.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends in cognitive sciences*, 5(10), 418-425.

Varela, F. J., & Bourgine, P. (Eds.). (1992). *Toward a practice of autonomous systems: Proceedings of the First European Conference on Artificial Life.* MIT press.

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind: Cognitive science and human experience.* MIT press.