

## SEEING RED: A POSTSCRIPT

Nicholas Humphrey

One day someone will write a book that explains consciousness. The book will put forward a theory that closes the “explanatory gap” between conscious experience and brain activity, by showing how a brain state *could in principle* amount to a state of consciousness. But it will do more. It will demonstrate just why this particular brain state *has to be* this particular experience. As Dan Lloyd puts it in his philosophical novel, *Radiant Cool*: “What we need is a transparent theory. One that, once you get it, you see that anything built like *this* will have *this* particular conscious experience.”<sup>1</sup>

Think of Einstein’s famous equation. The first step in explaining the equivalence of mass and energy is to recognize how energy, **e**, and mass, **m**, *could in principle* amount to the same thing when **m** is multiplied by velocity squared, **v**<sup>2</sup>, (because the two terms then have the same fundamental physical dimensions of mass, space and time).. But the second and all important step is to prove, with the theory of relativity, that **e has to equal mc**<sup>2</sup>, where **c** is specifically the velocity of light. So that, “once you get it, you see that anything with *this* mass will have to contain *this* particular amount of energy.”

In my book *Seeing Red* I put forward a theory of consciousness. Now, the question, which I hardly dare frame in these terms, is: does my theory of sensations as forms of bodily expression come anywhere near doing for the mind-brain identity equation what the theory of relativity did for the matter-energy equation? Does it take us any closer to understanding why a subject who engages in *this activity of redding in the brain* will have to experience *this particular red sensation*? Does it even pave the way to seeing how a theory *could* do this?

I am aware from early responses to *Seeing Red* that some readers – maybe you’re among them – while liking the book on other levels, think that with this big question I have

made no progress at all. So let me elaborate on why I think they're wrong.

I agree with Dan Lloyd that we want a *transparent* theory of consciousness. But, if we are not to beg the question, then, from the outset, we had better specify *transparent to whom*. In an essay I wrote for the *Oxford Companion to the Mind* (published in 2004, though written several years earlier), I set the bar high: I suggested we want a theory that is transparent not just to another conscious human being but to someone who is actually not conscious in the way that we are. I want to bring back this argument now to make my point.<sup>2</sup>

Lets stipulate, then, that the theory of consciousness has to be comprehensible to a scientist from Mars – an individual in many ways not unlike ourselves, highly intelligent, perceptive and even capable of self-reflection, but who nonetheless has never evolved into the kind of being who has sensations. Suppose we could explain to this Martian what happens in the brain of a human being who is engaged, say, in smelling a rose. And suppose he could thereby arrive at the entirely novel (to him) conclusion that it must *be like something to be this human being*, and indeed *like this* : “I am feeling this thick, sweet, olfactory sensation in my nostrils”. It's a tall order; but, still, it's what the theory ought to do.

Is a theory, which could bring this off, a possibility even in principle? Since the theory must employ only such concepts as the Martian can make sense of at the outset, we need to consider what kind of pre-theoretic notions he brings with him. Given that as yet he knows nothing about sensations, will he have other essential concepts on which to build?

We want him to understand that the human being is the *subject of sensations*. Can we assume he will at least have, to start with, the idea of what it is to be a “subject”? I'd say we can. For presumably the Martian is already himself a subject in the following crucial sense: an autonomous agent *who acts in the world*. Provided he can take himself as a model, he ought already to have the basic concept of an “I”. Then, can we assume he also understands the idea

of being the “subject of” something? Again, we can. For, as an “I” who does things with his body, he himself already has this genitive relationship to his own actions: he is the *author of everything he does*. So, will he even have the idea of being the subject of something with some of the peculiar properties of sensations: especially, that (i) they belong to the subject, (ii) they implicate part of his body, (iii) they are present tense, (iv) they have a qualitative modality, (v) their properties are phenomenally immediate? In fact he will: for analysis shows that *bodily actions already have precisely these characteristics (i) - (v)*.

Now, this may not seem much as a basis for understanding sensory consciousness. But I believe that, with the right theory, it will be enough. Suppose we suggest the following theory to the Martian (it is the theory I propose in *Seeing Red*, but others like it might also do the trick):

When a person smells a rose, he responds to what’s happening at his nostrils with a “virtual action pattern”: one of a set of action patterns that originated far back in evolutionary history as evaluative responses to various kinds of stimulation at the body surface – a form of bodily expression. In modern human beings these responses are still directed to the site of stimulation, and still retain vestiges of their original function and hedonic tone; but today, instead of carrying through into overt behaviour, they have become closed off within internal circuits in the brain; in fact the efferent signals now project only as far as sensory cortex, where they interact with the incoming signals from the sense organs to create, momentarily, a self-entangling, recursive, loop. The theory is that the person’s *sensation arises in the act of making the response* – as extended, by this recursion, into the “thick moment” of the conscious present; moreover, that the way he represents what’s happening to him *comes through monitoring how he is responding*.

Then, how will the Martian understand this? Presumably nothing in his own direct experience corresponds to what we have just described to him. But, still, he should be able to

work it out. He will be able to grasp the key fact that sensation begins as a kind of bodily expression, which is then self-monitored. He will be able to appreciate the peculiar temporal features of the feedback loop that has evolved. And so he'll be able to work out that *if* a subject like himself *were* to get involved in doing what the human being is doing, the result would be that he would have just *these beliefs about it, these attitudes, these things to say, these that he can't say*, and so on – in short he would experience it *like this*.

But if the Martian can work all this out from the theory, would this mean he actually acquires first-hand experience of sensations in the process? No: no more than someone who works out from physics and chemistry that  $H_2O$  constitutes water gets wet. A theory of consciousness is not a way of conferring consciousness; it is a way of understanding why consciousness-generating brain states have the effects on people's minds they do. In fact the Martian himself may have no sense organ with which to smell the rose at all: and yet, if the theory is right, he should still be able to discover *all* that we ourselves can discover by direct acquaintance. (And one day, of course, when *we humans* get to study Martians, the boot may be on the other foot).

Do you get it? Some readers of *Seeing Red* will still shake their heads. And I realize there is one piece of this argument that is especially hard to grasp. It is the bald assertion that for a person to be the subject of sensation simply *is* for him to be creating the appropriate sensory activity: for S to have a red sensation *is* for S to do the redding, to have the pain sensation *is* to do the paining, to have the salty sensation *is* to do the salting, and so on.

How *can* it be so? How can one thing simply *be* something else? The answer is that, according to the theory, it is *not* “something else”: the two terms – “doing the redding” and “having the red sensation” for example – pick out the same thing in the world.

Perhaps an example from modern physics may help make better sense of this. “String

Theory” posits that all the fundamental particles in nature are constituted by strings vibrating in nine-dimensional space. As Brian Greene explains:

Just as a violin string can vibrate in different patterns, each of which produces a different musical tone, the filaments of superstring theory can also vibrate in different patterns. But these vibrations don’t produce different musical notes; remarkably, the theory claims that they produce different particle properties. A tiny string vibrating in one pattern would have the mass and the electric charge of an electron; according to the theory, such a vibrating string would *be* what we have traditionally called an electron. A tiny string vibrating in a different pattern would have the requisite properties to identify it as a quark, a neutrino, or any other kind of particle.<sup>3</sup>

Thus String Theory claims that the vibrating string *is* an electron *because it has all the properties of an electron*. My theory of consciousness claims that a subject who is creating the appropriate activity in a reverberating circuit in his brain *is* having a red sensation *because he has all the properties of a subject having a red sensation*.

Well, perhaps both these ideas are hard to grasp. But it’s not clear that the latter is necessarily so much harder than the former.

I would not want to push the analogy with String Theory too far (there cannot of course be any real formal resemblance). But, now we have it in play, let’s see if this analogy can help with another more particular difficulty people have in understanding how consciousness can be constituted by brain activity. This is the problem of the qualitative gulf between sensations in different sensory modalities. How can one kind of activity in the brain be a sensation of color, while activity of another kind in the same brain is a sensation of taste, or of pain?

I suggest in Chapter Four of *Seeing Red* that the quality of sensations is determined by the “adverbial style” of the responses associated with different sense organs. And in an earlier book I drew my own musical analogy, and proposed that each type of sense organ has,

as it were, to be played like a musical instrument in its own way: “fingered, blown, bowed, plucked, etc., . . . so that the tactile modality might correspond to the woodwind style, the visual modality to the strings style, and so on.”<sup>4</sup>

Now, if you can see how, according to String Theory, “all species of particle arise from a different vibrational pattern executed by the same underlying entity”, perhaps you can also see how, according to my theory of consciousness, all qualitatively distinct varieties of sensation arise from a different type of reverberating activity in the same underlying nerve cells.

There we are, then. What’s it like to be an electron? I don’t deny that the hard problem of consciousness – the problem of how consciousness can *be* what consciousness *is* – is hard to get our minds round. We must seek help where we can get it. In *Seeing Red* I ask: What’s it like to be a painting, or a poem? Perhaps the poetry of superstrings can aid our intuitions, just as can the poetry of mist and bells.

And if my readers still can’t get it? Then maybe there really is something in the argument I put forward in the final pages of the book: that consciousness has been *designed* by natural selection to appear to us, its so-privileged subjects, to be just too hard to explain!

(Published online, March 2006)

1. Dan Lloyd, 2003, *Radiant Cool*, Cambridge Ma.: Bradford Books, p.16.
2. Nicholas Humphrey, 2004, “Thinking about feeling”, guest essay in *The Oxford Companion to the Mind*,. R.L.Gregory , ed., p. 213-4, Oxford: Oxford University Press. (Reprinted with permission; a few words have been changed) .
3. Brian Greene, 2004, *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*, New York: Knopf, p.18.
4. Nicholas Humphrey, 1992, *A History of the Mind*, p.165, London: Chatto and Windus.