The Hard Problem (but not that hard)

Mike Arnautov, 2019

After that splendid introduction by Tim, my task is to tackle the alleged Hard Problem itself, while other speakers will offer various alternative approaches to the issue of consciousness.

So what is “the Hard Problem”? The Stanford site has this to say (Van Gulick 2014: online):

> […] the so-called “hard problem” (Chalmers 1995) which is more or less that of giving an intelligible account that lets us see in an intuitively satisfying way how phenomenal or “what it's like” consciousness might arise from physical or neural processes in the brain.

That matches my understanding of the matter. It also raises a couple of big question marks. Firstly, and most obviously, why the demand foran explanation to be ‘intelligible’ and ‘intuitively satisfying’? Given that our scientific explorations of the world have led us to explanations which are experimentally verifiable and yet anything but intelligible or intuitively satisfying, such a demand does not seem at all reasonable.

And secondly, we have here an instance of a problem which bedevils all discussions on this subject. I suspect that other speakers won’t thank me for this admission, but I honestly do not think anybody discussing the matter (professional heavyweights included) really knows what they are talking about – quite literally, since there is no general agreement as to the meanings of basic terms being used.

Take 'physical or neural processes' giving rise to consciousness – what might that mean? It can't simply refer to processes as we understand them today. After all, our current theories of physics are certainly incomplete. Chalmers himself (1996: 95) seeks to avoid this difficulty in his book *The Conscious Mind* by appealing to ‘completed physics’, but what might that be? For all we know, it could encompass something which would strike us today as distinctly supernatural, because in the long run physicists incorporate into their theories anything that actually works, with no regard to earlier preconceptions. I don’t *think* we are likely to discover immortal souls, but who knows? There are aspects of our best current theories of physics, such as quantum fields for example, which Newton would almost certainly find deeply offensive to logic and intuition.

This being so, insofar as the Hard Problem is seen as an attack on physicalism, I do not feel to be in a position to defend physicalism – simply because I do not know what is actually being attacked. However, what I can do is to point out flaws in arguments advanced in defence of the Hard Problem, regardless of what those arguments are aimed to prove or disprove. We do not have enough time today for dealing with more than a couple of such arguments today – ones which seem to me most pertinent. Thus in the rest of this talk I’ll be addressing (a) the alleged problem of qualia and (b) the notorious zombie argument.

Let's begin by examining what would seem to be the most fundamental problem with the Hard Problem approach – the concept of qualia itself.

You probably do not need a reminder, but for completeness sake, qualia are generally understood as qualities of phenomenal experience – the 'what it is like' character of mental states; the introspectively accessible, specific ‘something’ making e.g. perception of red colour what it is (the quale of red, in this case). This is a bit vague, but far vaguer ‘definitions’ are also on offer. Can we do a bit better?

Eager to be educated, one turns to that font of all philosophical knowledge, the Stanford Encyclopaedia of Philosophy (Tye 2018: online), only to discover that (surprise!) there is no commonly accepted definition. Stanford gives four major ones. Qualia as phenomenal character. Qualia as properties of sense data. Qualia as intrinsic non-representational properties. Qualia as intrinsic, non-physical, ineffable properties. Needless to say there is much fine-grained argument and disagreement over these somewhat different conceptions of qualia. I do not propose to go into their details – we would be here for the rest of the day and then some. As I understand it, the exact nature of the beast is not supposed to affect the basic claim: qualia present a challenge to scientific investigation of consciousness. This challenge is usually presented in the form of thought experiments, so let us look at the one which is quoted most frequently in this context: the case of Mary, the shockingly mistreated colour scientist.

Mary's story, presented in its original form by Frank Jackson (1982: 42-49)[1], goes like this… From her birth, Mary is imprisoned in a black and white room. Never having been permitted to leave it, she acquires information about the world outside from black and white books her captors have make available to her, from black and white television sets attached to external cameras, and from black and white monitor screens hooked up to banks of computers. As time passes, Mary acquires more and more information about the physical aspects of colour and colour vision. Eventually, she becomes the world's leading authority on these matters. Indeed she comes to know *all* the physical facts pertinent to colours and colour vision. Then one day she is taken out of that room and seeing a red rose exclaims 'So that's what seeing red is like!'.

The point of the story being that despite knowing *all* there was to know scientifically about seeing colours, she still learned something new by looking at the red rose, so obviously, scientific knowledge cannot account for the phenomenal quality of the experience of red colour – R.I.P. physicalism.

There are, of course some problems with this tale. For example, one can object that it is far from simple to deprive one's victim of an experience of colours, but it is not hard to counter objections of this sort. Except here's an interesting one: how does one ensure that Mary does not dream in colour? Silly question, you may say: she has never **s***een* colour, so how can she dream in colour? And yet, I recently came across a report of a colour-blind man who also happens to be synaesthetic (Ramachadran and Hubbard 2003: 42-49). Though his eyes are unable to register colours, he experiences 'Martian colours' synaesthetically, which means that his brain somehow manufactures the general quale of colour (as opposed to of smell, touch,taste…) as well as the quale of a *particular* colour. Colour perception seems to be hard-wired into us, so perhaps one could dream in colour despite having no experience of colours.

As a separate issue, Mary is supposed to have learned *everything* there is to know scientifically about colour vision, but we know that perception of colour can be affected by neurological factors – from functioning of neural cells in one's retina, to the state of the neural machinery, which we know to be involved. So in order to have such complete knowledge, Mary would have to know the relevant neurological details of any given individual, herself included, otherwise her knowledge is purely abstract, with no link to the experience of any given individual

I submit that what Mary actually learns on seeing a red rose, is how the experience is registered by her specific visual apparatus and how it slots in with other experiences – because, as Husserl discovered in his investigations of phenomenology, even something as seemingly simple as seeing red colour is not an act of passive perception. Thus one can argue that Mary learns something new not about red colour as such, but about the way her specific neural apparatus responds to the sight of a red object.

It is, of course, still possible to insist that neurology is welcome to investigate the *origins* of qualia, but has nothing to say about their what-it-is-like phenomenal character. Why does the sensory experience of red presents itself to me with this vivid sense of redness? After all, none of the above explains the mystery of red looking… well, like red! To which the correct response, I think is: what *should* it look like? It is obviously useful to distinguish between red and other colours, otherwise evolution would not equip us with the neural machinery which allows us to do so. This means that the experience of red and of green must be different *somehow.* The quale of redness is simply your way of registering the red colour as opposed to any other experience.

You may object that I am evading an obvious problem: why does that experience of redness present itself to me as such vivid redness? Here we are at risk of slipping down the rabbit hole of dualism by separating the experience of redness from the way *it* presents itself to *me*. But there is no *it* separate from *me*. The quale of redness *is* our perceiving red colour – there is no 'presentation' involved. As to how it feels ('what it is like'), there is no way to compare what it is like for you with what it is like for me.. For all I know, if by some impossible magic I could experience your quale of redness, it would feel to me more like like some combination of texture and sound – perhaps furry in C major! It need not have any resemblance to my quale of red colour, other than fitting into the same overall structure of colour perceptions and of colour meanings – some wired into us, some learned.. Blood is red: red is danger, but red is also life. Red is 'warm', blue is 'cold'. Red is more unlike green than any other colour. All these 'overtones' blend together into one's unique experience of redness.

A defender of the Hard Problem could respond that while all this, hypothetically speaking, may or may not be true, none of it answers a deeper question: why should we be *conscious* of such qualitative specificity of experiences?

Of course, insofar as we do consciously experience anything at all, it seems obvious that our experiences must feature phenomenal qualities of some sort -- how do you consciously experience something without any such qualities? So it seems that the real question should be: why do we need consciousness?

This is, in fact a good question, which has nothing to do with qualia debates. For all we know, our consciousness is merely a consequence of some contingent turn taken by the evolutionary process. Does intelligence require consciousness or is it just an accidental feature of some terrestrial life forms? Could there be intelligent beings who lack consciousness? Unfortunately, in order even to consider the question we need to be able to say what we mean by 'consciousness'. And, guess what? – there is no agreement on that. Worse still, in the context of the Hard Problem, the question is asked very specifically in the context of human consciousness and we have no agreement on what that might be either.

This ambiguity around consciousness naturally leads us to the next major argument in the Hard Problem debate – the notorious philosophical zombie (or p-zombie) argument, which should have died soon after its birth (Chalmers 1996: 95). Alas, like a true zombie it marches on despite being killed over and over again. It goes like this: p-zombies, are beings indistinguishable from humans in all respects (physical, physiological and behavioural), except for lacking phenomenal consciousness (that ineffable something which makes life worth living, according to some). Such beings are imaginable. Therefore (?) they are possible. Therefore phenomenal consciousness cannot be accounted for by science. R.I.P. physicalism.

If that slide from 'imaginable' to 'possible' strikes you as dodgy, you are far from alone. To the best of my knowledge, no general defence has ever been offered for its legitimacy. Instead, defenders of p-zombies pick holes in any attempt to provide a counter-example of something being imaginable,

but not possible. Superficially, this is quite an effective strategy, but we must bear in mind that we routinely adjust our notion of what is imaginable, rather than just fanciful, in line what we think is possible. A cow jumping over the moon is imaginable to a small child but not, except in a purely fanciful way, to a scientifically informed adult. Thus our notion of what is really imaginable excludes what we know to be impossible, with no suggestion that everything imaginable could be actually possible.

There are, of course, other, well known objections. As Marvin Minsky observed (1998: online), the argument is circular. It *assumes* that physical characteristics of humans are not what produces phenomenal experiences and concludes that... physical characteristics of humans are not what produces phenomenal experiences. To drive the point home, Richard Brown (2010: 47-69) proposed an inversion of the zombie argument:

Let *zoombies* be beings *non*-physically identical to people in every way, except for lacking phenomenal consciousness. Such beings are imaginable, therefore (?) such beings are possible. Therefore physical aspects of human beings are quite sufficient for generating consciousness. Clearly, because of complete formal symmetry of p-zombies and zoombies nothing is proved either way.

The above should be sufficient, I think, to dispose of the p-zombie argument, but there is a further interesting problem. Suppose the p-zombie argument works. If it does, then p-zombies, being behaviourally indistinguishable from humans, would have to insist on experiencing the quale of redness (alongside of the rest of the qualia menagerie). They would praise that ineffable pleasure hit of the first (as opposed to the second and subsequent) mouthful of a cold hoppy lager on a hot day. They would admire the subtle interplay of colours in a beautiful sunset, they would be moved by the ephemeral beauty of an improvised jazz piano solo... So the question arises: how do we tell p-zombies from humans? Worse, how do we tell ourselves from p-zombies? I believe I have phenomenal consciousness – how is that different from *having* phenomenal consciousness, since, by definition, it is only accessible to me?

To be fair, Chalmers is aware of the problem. His response is (or was[2]) that while p-zombies may have such beliefs, their beliefs are necessarily false, because p-zombies, by definition, lack phenomenal consciousness. While their beliefs may have the same formal content as ours, and may result in behaviour indistinguishable from ours, there is this crucial difference – our beliefs on the subject of qualia are true, while theirs are false.

This seems an odd way of getting around the problem. Surely, consciousness is not just a matter of semantics. Nature does not care about our evaluative labels. The counter only works if one assumes *a priori* that consciousness, is just a human concept and does not feature in any possible 3rd person point of view ontology. We seem to be back to the sort of circular reasoning of the kind pointed out by Minsky.

There is, of course a very major problem lurking under the surface of all these arguments. Chalmers asserts that zombie beliefs must be false, because they lack phenomenal consciousness, whereas our otherwise identical beliefs are made true by our *direct experience* of phenomenal consciousness. The whole of The Hard Problem issue, presupposes that we do have such a direct experience and that this experience gives us a true picture of our internal world. It is the way it appears to us to be. But suppose that were not the case… I cannot help recalling Wittgenstein's alleged retort to the observation that it *feels* like the Sun going around the Earth rather than the Earth rotating: 'Well, what would it have looked like if it had looked as though the Earth was rotating?' To coin a phrase: phenomenology always under-determines ontology – a fact of which we should be very mindful.

The notion of consciousness as the active agent (or even the *only* active agent) within one's mind, as a kind-of Sun King of one's mental world, is so seductive, so pervasive, that even serious philosophers can taken in by it. But there are dissenters and I am happy to be counted as a very minor grunt in their ranks. In fact, I am prepared to meet Chalmers' challenge head on and say that, yes, in his specific terms I am a p-zombie. I still believe that I am conscious and have full access the phenomenal qualities of my experience. Those beliefs are (I believe) fully validated by my experience. If by a wave of a magic wand I was suddenly de-zombified, what would change, other than Chalmers proclaiming my beliefs flipping from falsehoods to truths, with no change to their actual contents or effects? And why should I care?

Which is all very well, but if consciousness is not the way it is traditionally thought to be, then what and how is it? Haven't we simply replaced one Hard Problem with another? Are there any promising alternatives on offer?

To my mind, the most promising direction is offered by horribly misnamed 'global workspace' theories of mind, originated by Baars (1988) and subsequently developed by others. Dennett's (1992) 'multiple drafts' theory can also be thought as belonging to this class, even though it was developed quite independently. Regrettably, this approach is not going to be covered by other talks, and I don't have the time to give you more than the barest sketch.

Baars' seminal suggestion was that consciousness should be thought of in terms of competition between non-conscious subsystems for a limited capacity, which broadcasts information for wide access and use. I called it 'horribly misnamed' for a good reason. The name suggests some special *place* within which consciousness somehow occurs – a throwback to some sort of the Cartesian Theatre, where non-conscious agencies report to the CEO of consciousness – a Trump-like figure, advised (and sometimes betrayed) but never constrained by his board of directors. That is a wrong image. A much better one, also used by Baars himself is that of a broadcasting system, making information within brain available on the to-whom-it-may-concern basis.

This chimes well with information management principles well known to software architects. Up to a certain degree of complexity of a large modular system, it is both feasible and efficient for modular sub-systems to communicate directly with each other. However, as complexity grows, eventually it becomes too costly to keep providing direct bi-directional links. An over-arching 'to whom it may concern' communication layer can be used instead to alleviate the problem and to provide a greater *ad hoc* flexibility. Since this does not require dismantling the older peer-to-peer communication links, the change can be evolutionary rather than revolutionary, which is clearly a bonus when trying to account for evolution of consciousness[3].

My personal favourite version of global workspace is the most radical one due to Oakley and Halligan (2017: online), which boldly proclaims that consciousness has no executive powers at all – it is merely a kind of virtual whiteboard, scribbled upon and attended to by non-conscious agencies, which do all the work traditionally associated with conscious thought.

To summarize, I have argued that the Hard Problem rests on shaky foundations. It demands intuitively satisfying explanations despite science routinely giving us counterintuitive yet accurate accounts of reality. It operates with poorly defined terms. And it assumes we have direct, accurate access to our own mental processes – that consciousness must be exactly as it appears to us.

In the meantime, new approaches to the matter of consciousness are being developed, which offer an alternative view of the matter. Yes, there is a lot of work to be done and the problem of consciousness is probably a genuinely the hard one, for reasons which have nothing to do with the Hard Problem.

Notes:

1. It should be noted that Jackson, the author of this particular intuition pump no longer supports its intended conclusion.

2. Quoted by Dennett in ‘Sweet Dreams’ (2006: 48). Dennett gives the reference as “Reply to Searle” on Chalmers’ website, but that reply does not appear to be there any more, hence I cannot claim that this represents Chalmers’ current views.

3. There is an interesting analogy here with Kahneman’s System1 (fast, reflexive, non-conscious) and System2 (slow, deliberative, conscious), as outlined in his “Thinking Fast and Slow” (Kahneman 2011).

References:

Baars B.J., (1988) A Cognitive Theory of Consciousness. NY: Cambridge University Press.

Brown R. (2010) ‘Deprioritizing the A Priori Arguments Against Physicalism’. Journal of Consciousness Studies, 17, No. 33-34.

Chalmers D. (1986) *The Conscious Mind.* NY: Oxford University Press.

Dennett D. (1992) *Consciousness Explained.* Penguin.

Dennett D (2006) *Sweet Dreams.* MIT Press.

Jackson F. (1982), ‘Epiphenomenal Qualia’, *Philosophical Quarterly 32*: 127–136.

Kahneman D. (2011) *Thinking, Fast and slow,* Farrar, Straus & Giroux.

Minsky M. (1998) ‘Consciousness Is a Big Suitcase’, *Edge.*
https://www.edge.org/3rd_culture/minsky/index.html

Oakley D.A. and HalliganP.W. (2017) ‘Chasing the Rainbow’. *Frontiers of Psychology 8:1924*

Ramachandran, V. S. and Hubbard, E. M. (2003) ‘Hearing Colors, Tasting Shapes’, *Scientific American*, 288: 42-49.

Tye, Michael (2018) ‘Qualia’, *The Stanford Encyclopedia of Philosophy.*
https://plato.stanford.edu/archives/sum2018/entries/qualia/

Van Gulick Robert (2014) ‘Consciousness’. *The Stanford Encyclopedia of Philosophy, https://plato.stanford.edu/archives/spr2018/entries/consciousness/*