

Nicholas Humphrey

A Riddle Written on the Brain

Abstract: *The sensation of red light falling on your eyes has something in common with the experience of looking at a cartoon in the New Yorker. The phenomenal quality of the sensation and the funniness of the joke are both properties of your subjective take on an external event and both arise in two steps. With sensations, (1) your brain responds to signals from bodily sense organs with an internalized evaluative response; (2) your mind reads this response and represents what it's like as the subjective property of redness. With jokes, (1) the cartoonist creates a clever drawing; (2) your mind takes in the drawing and represents what it's like as the subjective property of funniness. This analogy — deliberately deflationary — helps elucidate the nature of phenomenal consciousness and its neural correlates, and exposes the 'hard problem' as a conceptual error.*

Wittgenstein's parable about the 'beetle in the box' has left a long shadow. In my copy of *Philosophical Investigations*, I find I marked the page with a fishing permit dated 1962, when I was aged nineteen. I've been haunted by it ever since. You, like me, probably know the passage by heart:

Suppose everyone had a box with something in it: we call it a 'beetle'. No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. — Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. — But suppose the word 'beetle' had a use in these people's language? — If so it would not be used as the name of a thing? The thing in the box has no place in the language-game at all; not even as a something: for the box might

Correspondence:

Nicholas Humphrey, Darwin College, Cambridge. Email: humphrey@me.com

even be empty. — No, one can ‘divide through’ by the thing in the box; it cancels out, whatever it is. (Wittgenstein, 1958, §293)

Wittgenstein meant it to be primarily a cautionary tale about the problematical status of the words we use to refer to first-person experiences, especially bodily sensations. But many readers — certainly including me at the beginning — have taken it as calling into question whether sensations have any objective reality. If being the subject of private sensory experiences does not enable you to name them and talk about them, then what use is private experience at all in the public sphere? If sensations are ineffable, it’s only a short step to their being inconsequential, and if inconsequential then inessential. We are on the slippery slope that leads to epiphenomenalism, dualism, and philosophical zombies.

But what if Wittgenstein himself got off on the wrong foot? Consider an alternative scenario:

Suppose everyone had a slip of paper with a cartoon drawing on it that makes him laugh: we call it a ‘joke’. No one can look at anyone else’s slip of paper, and everyone says he knows what a joke is only by looking at his slip of paper. Here it would be quite possible for everyone to have slips of paper with different cartoons on them. One can even imagine the cartoon constantly changing. Now, suppose the word ‘joke’ had a use in these people’s language? If so, would it be used as the name of a thing? Well, sorry, Herr W., actually: yes. It would be the name for the cartoon that makes the subject laugh. The slip of paper cannot be blank! One cannot ‘divide through’ by the drawing on the paper, it does not cancel out, whatever it is.

Now the story works out rather differently, and the implications for the ontology of sensations must obviously be different too. But which way should we go? Are sensations more like beetles or like jokes? Fifty years on from my first impressionable encounter with the beetle — and having had considerable exposure to Daniel Dennett along the way — I must now say I think there’s no question sensations are more like jokes. Sensations do not cancel out. They are real ineliminable physical events inside our heads. Just as jokes have been designed by the cartoon artists to exhibit the quality of *funniness*, so sensations have been designed by evolution to exhibit the strange and beautiful qualities of *phenomenality* or ‘qualia’. If there’s a true analogy here — ‘phenomenal feel is to sensation as funniness is to jokes’ — we could be on to something. Indeed by looking at the workings of jokes we might actually learn about the workings of sensation.

So let's begin with some elementary questions: first about jokes and then the same about sensations. Imagine you open the *New Yorker*, and your eyes alight on a cartoon that depicts a fencer who has just sliced his opponent's head off exclaiming 'Touché'. Ha ha! That's a good one! But why do you laugh? Where is the funniness coming from? Clearly this question can be answered at several different levels. But I suggest we begin low down with the physical stuff involved. What can we say about the *material correlates of funniness*, the MCFs? Well, to start with, we have the actual drawing as it exists on the page of the *New Yorker*. Without this physical object there wouldn't be anything *to be* funny. We may say, then, the drawing is the *primary material correlate*, which we can call the MCF₁.

There must be more to it, of course. The joke doesn't end with the drawing on the paper. The drawing is not intrinsically funny. There's no way the page of the *New Yorker* will be enjoying the joke all by itself. The drawing does have some properties that could be said to be intrinsic: physical properties, for example — size, colour, the width of the lines, and so on. But funniness isn't one of these. Funniness is strictly a relational property that obtains between the drawing and you, the human observer. The drawing may be physically monochrome even when no one is looking at it, but it is funny — funny for you — only in so far as you represent it mentally as being so.

Mental representing — that's to say, taking in a state of affairs and attaching meaning to it — is a kind of computation that the human brain is particularly adept at. The mental content that results may seem to you, the subject, to be unrelated to and irreducible to any mere physical process (thinking something funny, as we'll see later, is not the half of it), but the representing is nonetheless something that is undertaken by a net of nerve cells. So when you *think* of the drawing as funny, thus creating a state of mind that is — psychologically — the cause of your laughing, the brain activity responsible must constitute the *secondary material correlate of funniness*, which we can call the MCF₂.

Now, let's ask the same questions about sensations. Imagine you see a brilliant sunset. Aha! How gloriously red the sun appears! But what lies behind the phenomenal quality of the sensation? Can we identify some part of the sensory process — starting with the sun, and ending with your experience of red qualia — that bears the same relation to the quality of this experience as a drawing on paper bears to funniness? That's to say, can we identify the primary material correlate of phenomenality, the MCP₁?

You might suppose the answer has to be the sun as such. Just as the drawing is the primary cause of your laughing, isn't the sun the primary cause of your having the sensation? Yes, it is in a way. But a moment's reflection will show that this can't be the right way to cut it. For, if the sun were to be playing the same role in sensation as the drawing is in the joke, we'd run up immediately against a major disanalogy. Where does *design* come in? In the case of jokes, the drawing has been created by a human cartoonist in order to make you laugh. That is what the drawing is meant for. But, in the case of sensations, it would be madness to suggest the sun has been designed by anybody to elicit red qualia, or that this is what the sun is meant for.

Could it be that the analogy is simply a bad one, and not worth pursuing after all? No, to the contrary, I think we should follow where it leads. If it's true and important that sensations are the product of evolutionary design, the designer can only have been natural selection acting on genes, and the equivalent of the drawing must lie where genes can control it: that's to say, within the brain. So in looking for the MCP₁, we should turn our attention to the chain of events that runs from stimulation of the sense organs to your having the sensation, and ask just where phenomenal quality could have been added.

I have my own ideas about this. For the sake of argument I'll outline them briefly. In writings going back twenty five years I've argued that the brain activity in question is an evolutionary ancient evaluative response to sensory stimulation — the response I've called 'sentition'. In the organisms that were our far distant ancestors, sentition originated as a set of reflex instinctive responses, forms of bodily expression, wriggles of acceptance or rejection. Having been honed by natural selection, these responses were precisely adapted to the situation, so that they took account of what kind of stimulus was reaching the body surface, what part of the body was affected, and what import this had for the organism's well-being. They took account of all this, and were indeed a physical reflection of it. So, from an outsider's point of view, sentition could be said to be carrying a wealth of information about the organism's sensory environs and how it evaluated them. Yet to start with the organism itself had none of this information 'in mind'.

I believe subjective sensation, as a mental state, first emerged when organisms evolved a way of accessing this information. The trick was for the organism to monitor the motor command signals being issued by its own brain. The command signals could then be read, in reverse,

as it were, to yield a picture of the stimulation the organism was receiving and how it mattered. In short sensation was — and still is in human beings today — the subjective reading of sentition: sentition taken to represent the experience of smell, colour, pain, and so on. And yet this subjective experience when it first evolved would not have had any special phenomenal feel. What happened?

I've argued that the key lies in how sentition went on evolving. In the early days it involved overt bodily behaviour. Eventually these overt responses ceased to be adaptive and were no longer wanted. However, by this time they had already come to have their new and useful role as the vehicle for subjective sensation. The answer to this adaptive dilemma was for the responses to become internalized or 'privatized', so that the command signals, rather than effecting changes in the actual body, began to target the internal body-map where the sense organs project to the brain. Thus sentition evolved to be a virtual form of bodily expression — yet still a response that the subject could milk for information about stimulation arriving at the body surface.

But the privatization of sentition had a remarkable, if fortuitous, result. It resulted in feedback loops being created between motor and sensory regions of the brain. These loops had the potential to sustain recursive activity. And, in the event, this development was game-changing. Crucially, it meant that, with changes to the gain in the loop, the activity could be drawn out in time, so as to create the 'thick moment' of sensory experience. But, more than that, the activity could be channelled and stabilized, so as to create a mathematically complex 'attractor' state. Such an attractor could in theory have remarkable hyper-dimensional properties. Real, unreal, magical? The answer might be in the eye of the beholder — the subject whose reading of this brain activity creates sensory experience.

At any rate, from now on, whenever the opportunity arose to 'improve' the quality of sensations — to make further adaptive changes — natural selection had a whole new design space to explore. Small adjustments to the circuitry could have dramatic effects on the subject's experience. This provided the evolutionary context, I believe, for the invention of a special kind of attractor state that would come to be experienced by the subject as a sensation with an unaccountable phenomenal feel. I've called this attractor state the 'ipsundrum', meaning self-generated conundrum. The ipsundrum is still a species of sentition, that originates as a response to sensory stimulation, and still carries information about the objective properties

of the stimulation. But this information now comes in a remarkable new package. It comes, if you like, as part of a riddle written on the brain.

Where has this got us, so far? We asked a few paragraphs back: what is your brain doing that bears the same relation to phenomenal feel as the cartoon drawing does to funniness? Here we have the update. It's the ipsundrum. To return to our example of watching a sunset, red light from the setting sun falls on your eyes, your brain makes an age-old evaluative response — we can call it 'redding' — and this gets worked up by feedback into an ipsundrum that invites interpretation as something with the feel of red qualia. Thus, in the terms we established earlier, the primary material correlate of phenomenality, the MCP_1 , is the brain activity that corresponds to the ipsundrum, and the secondary material correlate, MCP_2 , is the separate brain activity that monitors this and represents it as sunshine with phenomenal properties.

You do not of course have to buy into the details of the account just given. I've spelt it out as a proof of principle, to show how a materialist theory could possibly explain things. But if the details of the theory are negotiable, the bottom line is not. Sensation, as we've shown, has to be a two-stage process. Phenomenal feel — like funniness — arises only *with the representation*.

Now this bottom line may seem kind of obvious. And, when approached this way, I hope it is. But, I need hardly say it has not seemed obvious to everybody. In a well-known paper in 1995 Ned Block argued that we should distinguish two levels at which we are conscious of sensations: first 'phenomenal consciousness' and second 'access consciousness' (Block, 1995). This may sound like the distinction we've been making between the MCP_1 and the MCP_2 . And in a way it is. Block does believe there are two stages of brain activity, with phenomenal consciousness preceding and supplying the ground for access consciousness. The crucial difference, however, is that Block believes phenomenal consciousness — just as the name implies — already has a phenomenal feel before it has been accessed. Thus, quite contrary to the line we've taken, Block avers that phenomenality can be an intrinsic feature of brain activity. Indeed he has lately allied himself with the claim of neuroscientist Victor Lamme that cyclical activity in the temporal lobe of the brain is sufficient in itself to bring about phenomenal experience. In a 2015 paper titled 'Consciousness: Big Science and Conceptual Clarity', Block stoutly insists 'there can be conscious experience even without *actual* cognitive access' (Block,

2015, p. 5). But it's not just Block and Lamme. An unexpected number of contemporary philosophers seem to think — hope — this is the way to go. Michel Bitbol probably speaks for many of them: 'Can we preclude that any (large or small) area of the brain or even of the body is associated to some sort of fleeting pure experience, although no report can be obtained from it?' (Bitbol, 2014, p. 2).

There are of course a lot of things we can't preclude. But there are some we can. The notion of free-floating unaccessed phenomenal experience is conceptually opaque — about as easy to conceive of as a page of the *New Yorker* experiencing itself as funny. But even if we set aside the conceptual objection, the idea is empirically inert. By definition, no one could report or in any other way demonstrate that they were experiencing the phenomenal feel of a sensation unless they were to have some kind of access to it. So the claim lies beyond the possibility of scientific proof. Ironically, phenomenal consciousness, defined Block's way, would indeed be something 'we can divide through by', something that 'cancels out, whatever it is'.

I say the number of philosophers who side with Block is unexpected, because philosophers are meant to have a special talent for straight thinking. Their failure in this case is a grave let-down, not only for philosophy but for our wider intellectual culture. In some quarters it has brought the study of consciousness almost to its knees. Right and left — in books, on stage, in film — we hear talk of that notorious 'hard problem' of consciousness. What is this problem? It turns out to be nothing other than the problem of explaining intrinsic phenomenality: how a mere brain process could possess phenomenal qualities as a fundamental non-relational fact. Yet if, as I hope we've demonstrated, *no such thing arises in the natural world, nor ever will do*, the problem no longer exists. The philosophical effort put into solving it has been a colossal waste of time.

So, let's return to issues on which we *can* make further headway. When we diverted to Ned Block, our discussion of how sensations acquire their phenomenal feel was only half-way through. I'd argued that sentience has evolved to have such special features that, in the form of the ipsundrum, it invites representation as something with phenomenal feel. But if the ipsundrum has no intrinsic phenomenality, what exactly does it have to bring this off? How can a mere brain process seem to have phenomenal feel?

Well, what about this? I called the ipsundrum 'a riddle written on the brain'. In doing so I was continuing to allude to the analogy between sensations and jokes. But I wonder whether we can now take

this still further. Do sensations and jokes share structural features at the level of experience? If so, do they perhaps even share a mechanism? If we can discover how a mere drawing can seem funny, might we learn how a mere brain process can seem to have phenomenal feel?

The analogy is clearly not an identity. Sensations are not actually jokes. One thing they do have in common is that they both play — designedly — on your emotions. But they obviously do so in very different ways. Jokes make you laugh. Sensations fill you with wonder. Jokes make you write books with titles like *Inside Jokes: Using Humour to Reverse-Engineer the Mind* (Hurley, Dennett and Adams, 2011). Sensations make you write books with titles like *Soul Dust: The Magic of Consciousness* (Humphrey, 2011). You might think that in terms of their psychological impact the two classes of phenomena are just not in the same league. But, not so fast! Now it comes to it, I'd say we might quite well imagine a book called *Humour: The Magic of Jokes*. And, equally, a book called *Inside Sensations: Using Phenomenality to Reverse-Engineer Consciousness* (arguably that's just what we've been doing in this paper). So *can* the analogy deliver anything more? I think it's at least worth a try.

Hurley and co-authors (in the book just mentioned) have suggested a general theory of jokes. Other writers before them have observed, as they do, that the key to what makes a joke funny is that it sets up certain expectations in your mind, confounds them, and shows where you went wrong. However, few have gone further to ask just why funniness is *pleasurable*. Why should you *enjoy* being seduced into making a mistake and then having it corrected? Hurley *et al.* have an interesting and plausible answer: the reason you take pleasure in jokes — even if wicked pleasure — is that it encourages you to seek out further opportunities for testing and revising your own preconceived ideas. Jokes are needed to bump you out of the conceptual ruts that you've got into.

Jokes take you by surprise, and force a rethink. But we should note that this is as far as it goes. It's part of the deal that jokes do not shake the very foundations of your world-view. The authors point out that jokes are closely related to conjuring tricks (with the important difference that the conjuror doesn't show you where you went wrong). But we can note that the same house-rules as for jokes usually apply to tricks: you can take it for granted that stage magic is not *real magic*. And, if and when you do learn how the trick is done, you'll be able to

fit it comfortably into your existing picture of reality (and may indeed laugh out loud).

However, I want to draw attention to a third class of phenomena, that begin like jokes and conjuring tricks with mistaken assumptions, but end quite differently. These are what we might call ‘epiphanies’. Epiphanies occur when you are given evidence which, rather than requiring merely a rethink, forces you to revise some of your most basic tenets — to undergo a ‘paradigm shift’, in the terms of Thomas Kuhn. They lead you to embrace the magic rather than shuffle it aside. In such a case you may again feel pleasure because you’ve learned something new. But when what you’ve learned is indeed paradigm-shifting, you may feel an emotion on a different level — *awe*.

Epiphanies can arrive in many contexts: religious, romantic, philosophical, scientific. And they may be triggered by good evidence that a major revision is required, or bad. When John Taylor, a distinguished professor of physics in London, in 1974 witnessed Uri Geller bending a metal spoon by pure will-power, he was visibly knocked back. Shortly after, he wrote to Geller: ‘The Geller effect — of metal-bending — is clearly not brought about by fraud. It is so exceptional it presents a crucial challenge to modern science and could even destroy the latter if no explanation became available’ (Geller, 1975, p. 88). In this case, Taylor had mistaken a conjuring trick for the real thing. But his reaction was perfectly appropriate to what he *thought* he’d seen. On a different day it might have been Galileo responding to his discovery of the moons of Jupiter, Schrödinger to the outcome of the double-slit experiment, or Frege to the announcement of Russell’s Paradox (Frege wrote to Russell: ‘Arithmetic totters’).

Which brings me to sensations. How do you — how should you — respond to the mind-bending perplexities of conscious experience?

The record shows that even the best philosophical minds, when faced by what they *misconstrue* to be the hard problem, can surrender to awestruck babbling. Jerry Fodor is typical: ‘We can’t, as things stand now, so much as imagine the solution of the hard problem. The revisions of our concepts and theories that imagining a solution will eventually require are likely to be very deep and very unsettling... There is hardly anything that we may not have to cut loose from before the hard problem is through with us’ (Fodor, 2007, p. 9). Or Tom Nagel: ‘The existence of consciousness seems to imply that the physical description of the universe, in spite of its richness and explanatory power, is only part of the truth, and that the natural order

is far less austere than it would be if physics and chemistry accounted for everything' (Nagel, 2011, p. 35). These are, to say the least, interesting responses from professional philosophers to a phenomenon which, if we're right in our earlier analysis, is actually a magical effect that the brain lays on for itself — amazing, but no more paradigm-shifting than Geller's trick with the spoon.

But how about non-philosophers, ordinary people? What does your average Joe make of the fact that he seems to be the instantiation of a miracle? You might think most people are too down to earth to bother themselves with the meaning — the message — of consciousness. You might think so... until you look around you, and realize *there is no such person as a non-philosopher*. In fact the history of our species shows that human beings everywhere have evolved to be philosophers by nature, who are steeped in metaphysical anxieties, and as vulnerable as any professional — perhaps even more so — to the mistake of regarding consciousness as a paranormal revelation, *a handshake from another world*.

The big question then would be: could such a naturally contrived epiphany actually have had a positive outcome for human survival? Could it be that this is what consciousness, as a biological adaptation, is *meant for*? My book *Soul Dust* was devoted to showing that it could.

But I won't go there now. 'Suppose everyone had a box with something in it: we call it a "bloody marvel".' Enough said.

References

- Block, N. (1995) On a confusion about a function of consciousness, *Behavioural and Brain Sciences*, **18** (2), pp. 227–247.
- Block, N. (2015) Consciousness, big science and conceptual clarity, in Marcus, G. & Freeman, J. (eds.) *The Future of the Brain: Essays by the World's Leading Neuroscientists*, pp. 161–176, Princeton, NJ: Princeton University Press.
- Bitbol, M. (2014) On the primary nature of consciousness, in Capra, F. & Luigi, P.L. (eds.) *The Systems View of Life*, pp. 266–268, Cambridge: Cambridge University Press.
- Fodor, J. (2007) Headaches have themselves, *London Review of Books*, 24 June, p. 9.
- Geller, U. (1975) *My Story*, New York: Praeger.
- Humphrey, N. (2011) *Soul Dust: The Magic of Consciousness*, Princeton, NJ: Princeton University Press.
- Hurley, M.M., Dennett, D.C. & Adams, R.B. (2013) *Inside Jokes: Using Humour to Reverse-Engineer the Mind*, Cambridge, MA: MIT Press.
- Nagel, T. (2011) *Mind and Cosmos*, Oxford: Oxford University Press.
- Wittgenstein, L. (1958) *Philosophical Investigations*, Anscombe, G.E.M. (trans.), Oxford: Blackwell.