Topoi
DOI 10.1007/s11245-017-9498-0

The Invention of Consciousness

Nicholas Humphrey[1]

Abstract In English we use the word "invention" in two ways. First, to mean a new device or process developed by experimentation, and designed to fulfill a practical goal. Second, to mean a mental fabrication, especially a falsehood, designed to please or persuade. In this paper I argue that human consciousness is an invention in both respects. First, it is a cognitive faculty, evolved by natural selection, designed to help us make sense of ourselves and our surroundings. But then, second, it is a fantasy, conjured up by the brain, designed to change the value we place on our existence.

Keywords Consciousness · Evolution · Qualia · Illusionism

The literary critic William Empson said of his own profession: "Critics are of two sorts: those who merely relieve themselves against the flower of beauty, and those, less continent, who afterwards scratch it up. I myself, I must confess, aspire to the second of these classes; unexplained beauty arouses an irritation in me" (Empson 1930). We could say that students of consciousness are of two sorts also. On the one hand, those who want to see the mystery left intact, well watered but otherwise untouched. On the other, those who see it as a scientific challenge, a natural phenomenon that we need to dig up and explain.

Yet we all start from the same place. We relish the heat and redness of a fire, the sour tang of a lemon, the caress of a lover's hand. Mystic or sceptic, we all agree that consciousness is wonderful. Conscious sensations lie at the core of our being. Without them we'd be poorer creatures living in a duller world. What's more we all agree that consciousness is inexplicable—or at any rate that it is at present *unexplained*. The problem is not that we do not understand consciousness at all. Some aspects of it are relatively easy to account for in scientific terms. The problem is that *one* aspect continues to baffle everyone, and that's the "qualitative feel of consciousness": the redness of red, the painfulness of pain. The *qualia*—or, as people often express it, simply "what it's like."

The biologist H. Allen Orr probably speaks for the majority of scientists when, in a review of Thomas Nagel's book "Mind and Cosmos," he writes: "I share Nagel's sense of mystery here. Brains and neurons obviously have everything to do with consciousness but how such mere objects can give rise to the eerily different phenomenon of subjective experience seems utterly incomprehensible" (Orr 2013). Or, as Colin McGinn has colourfully put it: "The brain is just the wrong kind of thing to give birth to consciousness. You might as well assert that numbers emerge from biscuits or ethics from rhubarb" (McGinn 1993).

Well, let's see. I've called this paper "The Invention of Consciousness" because I want to play on two different meanings of the word "invention" in the English language.

An invention can be:

1. *A device or process, developed by experiment, designed to fulfill a practical goal.*

For example, a light-bulb or a telescope.

But alternatively, an invention can be:

Based on the "Mind and Brain Prize" lecture, Turin, Italy, 2015.

✉ Nicholas Humphrey
humphrey@me.com

[1] London School of Economics, London WC2A 2AE, UK

Published online: 09 August 2017

2. *A mental fabrication, especially a falsehood, designed to please or persuade*.

For example, a fairy tale or a piano sonata.

I am going to argue that human consciousness is an "invention" in both these senses.

That's to say, consciousness is:

1. *A cognitive faculty, evolved by natural selection, designed to help us make sense of ourselves and our surroundings*.

But, on another level, consciousness is:

2. *A fantasy, conjured up by the brain, designed to change how we value our existence*.

I'll argue that qualia make little if any contribution to the cognitive faculty. However they lie at the very heart of the fantasy.

I must start, of course, by defining the scope of the term "consciousness". People sometimes make a big meal of this. But I don't think this first step need be controversial—at any rate, not if we can ground it in the case we each know best subjectively, our own. If I may speak objectively on your behalf, consciousness is surely just *what you are conscious of*: that's to say the various states of mind of which at any one time you are the subject, and which are accessible to you by introspection.

It's true that consciousness, defined this way, may be difficult to access in nonverbal animals. But fortunately grown-up human beings can indeed *tell* us about it (at least up to a point). And what all agree is that you can be conscious of a range of rather different kinds of mental state: perceptions, memories, wishes, thoughts, feelings, and so on. When you introspect, you observe these various states, as it were with an inner eye. So, it comes naturally to you—and people everywhere do this—to think of consciousness as some kind of window on the mind, a private view of the stage where your mental life is being played out.

A view from whose standpoint? Well, from the standpoint of whom else but "you", your *self*. And this brings us immediately to one of the most striking features of consciousness: its *unity*. There's only one "you" at the window. Only one self. When you find yourself feeling pain, or wanting breakfast, or remembering your mother's face, it's the same you in each case.

We might think it obvious that it has to be so. But actually this unity is by no means a logical necessity. I'd say it's quite conceivable—and indeed psychologically plausible—that your brain could house several independent yous, each representing a different segment of the mind. Indeed this fragmented state may have been the way you and every other human being started out at birth. Back then, and for the first few months of life, the different yous might hardly have known each other. Thankfully, however, it was never going to stay that way. As your life got going and your body—your *one* body—began interacting with the outside world, these separate selves were destined to come into register—orchestrated, as it were, by the single line of music that, as it happened, made up your one life (Humphrey 2000).

Was this "binding of selves" genetically pre-programmed? Not necessarily. I think it could have been the automatic outcome of the dynamics of mind and body. In fact, something like it can be seen occurring in quite simple physical systems. In the seventeenth century Christian Huygens, the inventor of the pendulum clock, made a surprising observation. When two or more of his clocks were hung from the same beam, he noticed that their pendulums spontaneously began to beat in synchrony, showing as he put it an "odd kind of sympathy". In a more recent demonstration, a set of five metronomes are placed on a floating table, and they too soon begin beating as one (Harvard Natural Sciences Demonstrations 2016). It happens because each individual metronome, interacting via the table, feels the pull of the others. In the case of consciousness, presumably the story must be more complex. Yet perhaps not very much more complex. Perhaps the separate parts of a newborn mind, interacting with a single body, also somehow feel the pull of the others.

Whatever the truth of this, let's turn to the big question. Once your mental states all have the same subject, what does this unity achieve? The answer is a big one too. The unity of consciousness underwrites the most obvious cognitive function of consciousness, which is to create what Marvin Minsky has called the "society of mind" (Minsky 1986). Just as—in fact just because—there is only one "you" at the window, there comes to be only one mind on the other side. Information from different agencies is being brought to the same table, as it were, and it's here that your sub-selves can meet up, shake hands and engage in fertile cross-talk. This means you now have a mind-wide forum for planning and decision making. And the way is then open for a *central processing unit* to take control: an intelligent agent that can recognise patterns, marry past and future, assign priorities and so on across the mind as a whole. A computer engineer might recognise this as an "expert system". You of course recognise it as "I".

But, alongside this, another opportunity emerges. Once you can observe the parts of the mind interacting on a single stage, you are in a position to *make sense* of the interaction. And this can support a second important function of consciousness: namely, to allow you to appreciate just *how* your mind works. Observing, for example, how "beliefs" and "desires" generate "wishes" that lead to "actions",

you find your mind revealed as having a clear psychological structure. Thus you begin to gain insight into *why* you think and act the way you do. This means you can explain yourself to yourself, and explain yourself to other people too. But, equally important, it means you have a model for explaining other people to yourself. When you meet another person, you can assume his mind works much as yours does. So you can work out what he is likely to be thinking and how he will behave. Consciousness has laid the ground for what psychologists call "Theory of Mind."

So far, so good. We have a workable definition of consciousness in terms of introspection. And we've identified two ways in which introspection can be put to practical use. So that's two reasons why this kind of consciousness would have been likely to be selected in the course of evolution. What's more we have a plausible metaphor for how it works: consciousness provides a window onto—and at the same time creates—the society of mind.

Yet, what about the imagery I'm using here? Doesn't it smack of the "Cartesian theatre" on which Dan Dennett (1991) has poured such scorn? No, I think that's a false worry. What Dennett has objected to is the idea that the brain contains a projection space where a replica of the outside world is on show to an inner observer. But I hope it's clear this is not what's being proposed. What the window of consciousness opens onto is a picture not of what's outside but of what's inside—the mental states whose turns and twists and conflicts underlie the way you think and act. If this is theatre, it is indeed more like a proper human theatre, where a *play* is running.

Imagine yourself at a performance of a Shakespeare play. Shakespeare was not concerned with copying reality. His plays are stories, dramatic mock-ups, designed to analyse, expose and explain. And indeed as he himself made plain, the stories rely on codes and shorthand. In a famous prologue to Henry V, the Chorus apologises on behalf of the actors—mere *ciphers or symbols*—for daring to recreate the pageant of history on stage. "Pardon," the Chorus says, "The flat unraised spirits that have dared on this unworthy scaffold to bring forth so great an object: Can this cockpit hold the vasty fields of France?" The secret, he continues, lies in the encryption. Just as a string of zeros can represent a huge number—"Since a crooked figure may attest in little place a million"—so the players and props on stage can represent a reality of quite a different order. "So let us, ciphers to this great account, on your imaginary forces work."

It's a startlingly prescient passage—almost as if Shakespeare has anticipated modern ideas about how mental states are represented in the brain. But, with his words in front of us, I want to take up another remarkable allusion: "Can this *cockpit* hold the vasty fields of France?"

The term "cockpit" originated of course as the name of an arena for staging cock-fights. Already, by Shakespeare's time, it had morphed into the name for any confined space where important things get done. He could not have known that the word cockpit would later come to mean the wheelroom of a ship and later still the control room of an aeroplane. Yet, now, when we're discussing consciousness, I want to suggest the cockpit of a plane provides an even better analogy for consciousness than the theatrical stage does.

So picture, if you will, the cockpit of a plane. And place yourself where the pilot sits. You'll see before you an array of instrument panels, that display the output of a variety of modules that are monitoring the plane's external and internal states: speed, altitude, fuel reserves, global position, intended course, and so on. Let's say then that, from your privileged seat, you have a window on the plane's beliefs, desires, and intentions—presented in coded form, of course, as numbers, icons, graphs. Your job as pilot is to integrate all this information, so as to decide what to do to achieve certain goals. You must observe, then think, then act. You have a joystick with which you can control the plane's wing flaps and tail fin, so as to steer the plane in the intended direction. Oh, and by the way, you also have a cockpit radio, so you can report verbally to ground control. You have become in effect the *plane's self*.

You'll appreciate the analogy. And yet, you may be wondering what the point is. A *conscious human pilot* as an analogy for a *conscious agent* in the brain? If there's consciousness on both sides of the equation, where does that get us? But that's just it. It doesn't have to be on both sides. I want to use the analogy as a further way of demystifying consciousness.

We already know for a fact that there's no need to have a conscious agent in the pilot's seat. An electronic autopilot, made of nothing but circuit boards, can—and in many planes does—fulfill exactly the same function as the pilot, collating information, referencing a knowledge base, choosing the best path, and so on. The autopilot can even be designed to report on what it's doing and why, to a base on the ground, in simulated speech if required. And it can keep a historical record of its own activity (tucked away in a black box so that it can be accessed posthumously if necessary).

True, no one has yet engineered a plane's autopilot to be capable of reading the minds of other planes. But as it happens just such meta-cognitive abilities are already being incorporated into the computers of driverless cars. To navigate traffic safely, the computer must be able to anticipate how other cars are likely to behave. The computer has to have, in effect, a "Theory of Drivers". How does it learn this theory? I don't know the facts here, but I wouldn't be surprised if engineers are already working on having one

computer learn how to model other computers by reflecting on its own example.

So, back to the problem of consciousness. My point of course is that if an electronic autopilot can be engineered to do all this, then it's not so surprising that a brain can. We're talking normal science and engineering here. In fact the science is well under way. To mention a few areas of good progress: Stanislas Dehaene (2014) has been mapping what he calls "the global neuronal workspace"; Giulio Tononi (2012) has proposed a statistical model of "integrated information"; Christof Koch and Francis Crick (2005) have identified a brain structure, the claustrum, as a potential candidate for the master of ceremonies.

I suggested at the start that consciousness is an invention in the first sense of the term: "A cognitive faculty, evolved by natural selection, designed to help us make sense of ourselves and our surroundings". Exactly. So far it seems this is just what consciousness is. And, as I suggested would be the case, we haven't yet had to say anything about the mysterious *feel*. We get this cognitive faculty—the workspace, the integration, the theory of mind—without having so much as to mention the *eeriness* of consciousness.

This is good news, in its way. But bad news too. The good news is that we're getting an account of consciousness that looks like being scientifically respectable. The bad news is we're getting an account of consciousness that leaves out the very thing that many of us think of as its most baffling and intriguing feature. What about the eery phenomenal feel of consciousness? Where's the "what it's like" that everyone beefs about?

We defined consciousness at the outset as comprising all those mental states that are available to introspection. But now, if we want to make the eeriness of consciousness the issue, we'll have to focus in. Does the quality in question pervade all mental states? No, that's the thing: it does not seem to be a feature of higher-level cognitive states. At any rate it's not a necessary feature. There is no special *feel* associated with your having the thought, say, that today is Thursday. It's not *like anything* for you to believe it's going to rain, or to remember where you put your hat.

Rather, it seems the phenomenal quality kicks in only at a more animal level. It's there especially, perhaps exclusively, in the way you represent what's happening at your bodily sense organs—skin, eyes, nose, ears, tongue. It's there—and it's only there—with your experience of *sensations*: the pain of a bee sting, the salt taste of an anchovy, the blue look of the sky. Among conscious mental states, sensations have the very special property of being *intrinsically eery*, they simply couldn't be the states they are without having this mysterious dimension to them.

As I said at the opening, sensations lie at the heart of our being. No one would or could wish *qualia* out of existence. Indeed there will have been times for all of us when conscious experience is *about* little else. A science of consciousness that leaves qualia out is not just ignoring the elephant in the room, it is ignoring the elephant that *is* the room. Yet so far it seems that this is all the science we're getting. How can that be?

There may be several explanations for why qualia are not been given the priority we might expect. No doubt it's partly because, as we have just seen, cognitive science can indeed go a long way towards explaining consciousness without any reference to them. But it's also because of the fear, expressed by a good many scientists—and philosophers too—that it will never be possible to explain qualia in conventional scientific terms. H. Allen Orr, as we saw, said that qualia are "utterly incomprehensible". Christof Koch wrote to me not long ago: "it is bizarre that brain matter should exude these phenomenal feelings. Consciousness is so vivid, and its properties appear so otherworldly, that it seems to call for God". Koch may have been half-joking. But who's laughing? Short of invoking some supernatural agency, where are we to go?

There are indeed a good many theorists who simply don't want to go anywhere with it. It's not so much a case of *qualia denial*—though that exists too—as *qualia avoidance*. Isaac Newton set the tone 500 years ago: "But, to determine more absolutely, what light is, after what manner refracted, and by what modes or actions it produceth in our minds the Phantasms of Colours, is not so easie. And I shall not mingle conjectures with certainties" (Newton 1671). Jerry Fodor has echoed Newton's pessimism: "We don't know, even to a first glimmer, how a brain (or anything else that is physical) could manage to be a locus of phenomenal experience. This is, surely, among the ultimate metaphysical mysteries; don't bet on anybody ever solving it" (Fodor 1998).

Of course not everyone has been so ready to surrender. In the coffee room, if not yet the lab, there has been ongoing debate about just what kind of thing qualia are and what to do about them. The answers that have been proposed have not always been helpful. Yet it does seem a consensus is emerging, at least about the boundaries of the problem. Most theorists now accept that there are only two options that can be taken seriously. We can be *Realists* about qualia, or else we have to be *Illusionists* (Frankish 2016).

The names make the meaning of these alternatives clear. Realists take qualia at face value. In their view, if your sensations appear to have qualities that lie beyond the scope of physical explanation, then it must be they really do have such qualities. And this is possible because the brain activity that underlies sensations already has consciousness latent in it as an additional property of matter—a property as yet unrecognised by physics, but one that you the conscious subject are somehow able to

tap into. Tom Nagel, for example, writes: "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 2012). So, according to the Realists, when you experience pain, say, you are in effect breaking through the veil of mundane physics to access a higher-order realm.

Illusionists, by contrast, will have none of this. They argue that if your sensations appear to have these marvellous non-physical properties, then this can only be because your physical brain is playing tricks on you. And this is possible because the brain is a computational engine that deals in symbols, and physically based symbols can perfectly well represent states of affairs that do not and even could not exist (thank you, Shakespeare!). Dan Dennett, for example, has it that: "Consciousness is an illusion of the brain, for the brain, by the brain." Qualia are like "a beautiful discussion of purple, just about a colour, without itself being coloured" (Dennett 1991, p. 371). So, according to Illusionists, when you have a sensation—of purple, or sweetness, or pain—you are accessing your own brain's magic show and being *tricked into believing* you have reached through to another level of reality, when in fact it's all coming from your side.

Realism and Illusionism. The trouble is that both these theoretical positions come at a considerable price. On the one hand, the price of Realism is that it implies that the standard physical description of the world is radically incomplete. Some people actually welcome this. Nagel thinks it would make the natural order less austere! But others—including me—find it a lazy and inelegant solution.

But then, on the other hand, there's a price to illusionism too. Illusionism undermines—and in many people's eyes devalues—the mystery of human experience. Some people welcome that too. Dennett clearly takes wicked delight, in discomforting what he calls the Mysterians. He's happy to be, as he puts it, "the cop at Woodstock" (the policeman at a pop festival). But many others find illusionism deeply depressing, complaining that it "unweaves the rainbow" and so on.

Still, which is right? No one yet knows for sure. But I'm not hiding which I hope is right. Although I myself have recently questioned the language of illusionism (Humphrey 2016b), I hope to see a resolution of the "hard problem" within the bounds of our standard world model.

Here's an appealing analogy. I expect you are familiar with the "real impossible triangle", or "Gregundrum", a wooden object invented by Richard Gregory which, when looked at from one particular viewpoint, looks exactly like a solid Penrose triangle—a structure that simply couldn't exist in the physical world. My suggestion—my hope—is that the apparent "unreality" of consciousness comes down to a similar trick of perspective.

Can we do better than merely hope for this? Does anyone have any idea about what kind of physical processes in the brain might possibly underlie it? Actually yes, as I'll explain in a moment, I think—contrary to Fodor—we do have at least "a first glimmer". But before going there I want to consider a much simpler example. When sceptics are questioning whether any scientific theory can deliver the semi-magical effects, it will be good if we can point to a model mechanism that can emulate some of these effects. Then, at least we'll have a proof of principle.

So let's go back to my cockpit analogy. And let's suppose now that the plane you are flying has specialised sensors in its body, analogous to human sense organs, whose job is to represent what's happening at its body surface—heat, pressure, tissue damage and so on. Let's suppose, too, that there is a special set of "sensory instruments" in the cockpit, which display this information. But here's what's special: while all the other instruments on the panel use simple flat graphical or numerical displays, the sensory instruments—and only the sensory instruments—dress them up in a very special way: as *holograms*.

We've all seen holograms. The picture appears to rise above the flat surface. Of course we know it's not real. It only looks as if there's a third dimension. However, you, in the magical cockpit *don't know this*. To *you* it seems that the numbers really are jumping out of the screen. No wonder, then, that you find these sensory displays specially attention-grabbing and impressive. You do your best to explain to others, over the radio, just what it's like. But sadly, words often fail you. Still, it is your own first-person experience that matters to you above all. From now on you will go flying just to immerse yourself in these extraordinary displays. As Lord Byron said: The great object of life becomes sensation—"to feel that we exist, even though in pain" (Byron 1813).

But I must not get carried away, just because you the pilot have been. I'm running ahead of my own argument.

OK. An analogy is an analogy. A hologram is a hologram. What can this actually have to do with the brain and qualia? Well, dare I say it, maybe it's not just an analogy. I want to draw your attention to the so-called "holographic principle" which has come out of cosmology and the physics of black holes. The principle states that, not only can a three-dimensional world always be represented without loss of information by a two-dimensional surface (as in a conventional hologram), but *an n-dimensional world can always be represented by a (n – 1) dimensional surface*.

Thus, to start with, when three-dimensional objects disappear into black holes, the information they contain need not have been finally lost—which would be problematic for physics—but instead could be preserved on the hole's

two-dimensional surface, *from which an illusion of the original objects could be regenerated*. In fact, in light of this, cosmologists have suggested that the three-dimensional world we ourselves believe we inhabit could actually be just such an illusion arising from a flat two-dimensional surface. But more to the point, we can now suggest that the four-dimensional world of conscious qualia could quite well be an illusion generated by a three-dimensional brain. As someone said about the black hole case: "This idea is so odd, it's comparable to finding that the instruction manual for a dishwasher holds the recipe to making a good chocolate soufflé" (Maynard 2015). Ah ha! As someone else said about consciousness: "You might as well assert that numbers emerge from biscuits or ethics from rhubarb"(McGinn 1993). Looks as though we might be on to something!

Yes, but how precisely could it work? As it happens, Karl Pribram, back in the 1970s, did indeed raise the possibility that information in the brain is stored in holograms. But no one today takes Pribram's detailed model seriously. So how else might the brain be generating a higher-dimensional sensory display? I've been working on an answer to this question for many years (Humphrey 1992, 2006, 2011). I've wanted an answer that takes account of evolutionary history. This isn't the place to give you the full story, but I'll try to give a brief overview.

It begins, as I see it, with the creatures that were our far distant ancestors, floating in the seas, making evaluative responses to stimuli at the body surface: "wriggles of acceptance or rejection". These responses, to which I've given the general name "sentition", have been honed by natural selection, so as to be well adapted to the creature's needs—taking account of what kind of stimulus is reaching the body surface, what part of the body is affected, and what import this has for biological well-being. From the start then, the responses can be said to be *meaningful*—which is to say they potentially carry a lot of information about what the stimulation means for the creature. However, to begin with, there is no one at home in the brain to realise this potential, no one to *take an interest* in the meaning.

But, evolution is inventive. Before long there arises in the brain a special module—a proto self, if you like—whose job is exactly that: to discover "what the stimulation means for me". And, as luck would have it, it turns out it can do this by the simple trick of *reading—extracting the meaning from—the motor command signals* being sent out to produce the reflex response.

So now, we have an agent who is reading the brain's own responses and making a sensory interpretation of them. In truth this is the first *subject of sensation*. But let's note there is nothing fancy or magical about the interpretation at this stage. The subjective experience does not have had any special phenomenal feel. What happened?

I've argued that the key lay in how sentition went on evolving. Back at the start, the reflex responses are overt bodily actions occurring at the site of stimulation at the body surface. However things are never going to stay like this. As the descendants of the original creatures evolve to be more sophisticated, these overt responses soon enough become inappropriate, even inconvenient—you don't always want to grimace when you're touched by red light, say. So now the creature faces a problem. How to lose the bodily behaviour but keep the information about the meaning of the stimulus?

The solution natural selection hits on is ingenious. It is for the responses to become internalised, or "privatised", such that the motor signals no longer reach the actual body surface, but rather begin to target the body-map where the sense organs first project to the brain. Thus sentition evolves from being an actual form of bodily expression to being a *virtual* one—yet still a response that the subject can milk for information.

Now, this privatisation has a remarkable—if fortuitous—result. It means that a feedback loop is created between motor and sensory regions of the brain—a loop that has the capacity to sustain recursive activity, going round and round, catching its own tail. And this, as I see it, is game-changing. Crucially, it means that the activity can be drawn out in time, so as to create the "thick moment" of sensory experience. But, more than this, the activity can be channelled and stabilised, so as to create a mathematically complex "attractor" state. And such an attractor can have remarkable hyper-dimensional properties (Krisztin 2008). Real, unreal, surreal? The answer will be in the eye of the beholder—the subject whose reading of this brain activity is giving rise to the sensory experience.

At any rate, from now on, whenever the opportunity arises to "improve" the quality of sensations—to make further adaptive changes—natural selection has a whole new design space to explore. Small adjustments to the circuitry can have dramatic effects. And this provides the evolutionary context, I believe, for the invention of a special kind of attractor that will be read by the subject as a sensation with an unaccountable *phenomenal feel*. On the analogy of the Gregundrum, I've called this attractor the "ipsundrum", to signify a real "impossible brain state" that is actually self made. 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 guise. It comes, if you like, as part of "a riddle written on the brain" (Humphrey 2016a).

I put forward this account of sensations more than 25 years ago. My arguments were largely theoretical, rather than empirical. But I'm happy to say it looks as if one of the key features has been getting experimental backing:

namely that visual sensations depend on brain activity in a loop running between primary visual cortex and areas further forward. In a masterly review of recent neuroscientific evidence, Stan Dehaene (who, oddly enough, is something of a "qualia denier") sums up the picture he sees emerging: "Consciousness lives in the loops: reverberating neuronal activity, circulating in the web of our cortical connections, causes our conscious experiences" (Dehaene 2014, p. 156).

So there we have it: my glimmer of a theory of what gives consciousness its astonishing quality. With so much of the detail missing, I acknowledge it's not much more than a glimmer. But it must be better than no theory at all. Colin McGinn has written: "It is not that we know what would explain consciousness but are having trouble finding the evidence to select one explanation over the others; rather, we have no idea what an explanation of consciousness would even look like" (McGinn 1999, p. 61.) I humbly suggest that's no longer true.

This is all I have to say for now about how a physical system could deliver conscious experience. However, for an evolutionist, of course it's too soon to wrap up the discussion. We may have found a possible answer to the question of *what* evolved, but we haven't yet begun to address the question of *why* it evolved. Even if we did know all the detail—if we could explain how conscious experience is created neuron by neuron, from red light touching your retina through to your making all the claims you do about the red qualia—we still would not know *what this is good for*. What can possibly have been the biological advantage, the contribution to fitness, of dressing up sensations in this provocatively mysterious way?

It's a real problem. Let's return to the idea of consciousness as an invention. Under the first meaning of invention we saw that consciousness could indeed be considered to be *"*a cognitive faculty, evolved by natural selection, designed to help us make sense of ourselves and our surroundings.*"* But now, when we consider the role of qualia, this meaning of invention looks much less of a good fit. At first sight at least, *qualia* are neither cognitive, nor helpful!

Jerry Fodor has stated the difficulty in his typically blunt way: "Consciousness"—and it's clear he's referring to qualia in particular—"seems to be among the chronically unemployed. As far as anybody knows, anything that our conscious minds can do they could do just as well if they weren't conscious. Why then did God bother to make consciousness?" (Fodor 2004). John Searle has made much the same claim, about qualia having no impact at the level of behaviour: "As far as the ontology of consciousness is concerned, behaviour is simply irrelevant. We could have identical behaviour in two different systems, one of which is conscious and the other totally unconscious*"* (Searle 1992).

If these philosophers are right, it would mean that consciousness—at least its phenomenal side—could not have had any impact on our ancestors' survival. In which case the genes specifying the underlying brain circuits could not have been selected by natural selection.

Then, *are* these philosophers right? I think the plain answer is, No. They are guilty of a massive failure of imagination.

Fodor says qualia are "unemployed". He seems to take it for granted that, if consciousness does have a job to do, this can only be to provide us with some special kind of skill—helping us to act more intelligently or more efficiently in the service of some practical goal. But what if this notion of employment is simply not appropriate when discussing the phenomenal aspect of consciousness? What if phenomenal consciousness, rather than making us more intelligent or more productive *on the outside*, makes us somehow *bigger on the inside*—emotionally and spiritually bigger? What if consciousness is actually an invention in the second sense I mentioned at the start: "a fantasy, conjured up by the brain, designed to change how we value what becomes of us?"

Think about it. Think again about the real impossible triangle, the Gregundrum. Why, for what purpose, did Richard Gregory invent this brilliant illusion? It surely wasn't to serve any practical purpose. There's a photo showing him with his face framed by the real impossible object (Gregory 2011). Look at his broad smile. He did it simply to *amaze us*. Then, could it be that Nature, when she invented qualia, did it so that we conscious creatures should *amaze ourselves*?

Don't get me wrong. I am a card-carrying Darwinian reductionist. I've no wish to get off the explanatory hook by substituting fuzzy answers for clear ones. But still, I do think there are times when, in the interests of science, we need to loosen up a bit. Before we pronounce on the employability of phenomenal consciousness, we need to undertake a proper natural history. We should be studying how conscious experience actually changes the way people live in the world. How does exposure to qualia change people's psychology? What beliefs and attitudes are generated? How does it affect people's ideas about who and what they are, and what kind of world they live in?

These are—or ought to be—empirical questions to be asked of ordinary people. And we should be ready to consider all sorts of possible answers, not just those we'd find discussed in the science or philosophy section of the library but perhaps those that belong in the self-help section, or even the New Age. But, most important, we should begin the inquiry close to home, by taking seriously our own intuitions about just how and why phenomenal consciousness matters to ourselves.

Think about it. Suppose the magic for *you* were *not* there. Suppose your sensations were in fact just brown bag numbers. What would be missing from your life?

It's clear to me that in such a semi-zombie state I—you—would lose out, on several levels. First, you'd lose your psychological essence, your core self. Next, you'd lose your sense of intimacy with things in the outside world. And then, finally, you'd lose your soul, and other humans would lose their souls as well.

1 Self

We saw, early on, how the binding of sub-selves leads to the creation of the core self as the singular subject of a range of mental states. But, now let me say it, even when all the sub-selves are gathered together, the larger self is by no means secure. A self stripped of sensations would remain a pretty anaemic kind of self. But add in the qualia, and everything changes. By lifting sensory experience onto that mysterious, non-physical plane, qualia deepen and enrich your sense of your own presence. You find yourself living in thick time. You become the owner of a self that you want to expand and preserve for its own sake—in short, a *self worth having*. Take away this primary sense of your own presence, and your existence would simply be less well-founded, less convincing—to you and everyone else.

2 World

Next, though this isn't so obvious, you'd lose the external world—at least the world as you've come to know and love it. Even though it's your own brain that creates the qualia, you can't but project the special qualities of sensations out onto the objects of perception in the outside world. In doing so, you spread a kind of fairy-dust around you. You enchant the world. Take away this magic paintbrush, and the world would lose much of its significance. You'd find it a less awesome place, less fun, less promising.

3 Soul

You did it. It's all yours. The things out there, experienced through bodily sensation, are singing your song. It's bound to dawn on you that when you pay homage to the beauties of nature you are really paying homage to yourself. So, by a strange inversion, the magical world you've made returns the compliment and further enhances your sense of your own significance. Then add in the poetry of human culture, and by one path or another, your core self becomes elaborated into that marvellous cultural construct: the human soul. A soul that, with your generous theory of mind, you recognise in other people too.

Now, I *will* draw this to a close. Earlier, when I quoted Shakespeare's prologue, I omitted the first lines. They read.

> O for a Muse of fire, that would ascend.
> The brightest heaven of invention.

The chorus means "invention" in the second sense: he's seeking permission for the actors to create an extraordinary work of fiction on the stage. I like to think that Nature did it first. Qualia are just such an invention, arguably the brightest heaven—the most remarkable story that anyone has ever dared to tell. Thanks to natural selection, we all contain within ourselves that muse of fire.

Compliance with Ethical Standards

Conflict of interest The author declares that there is no conflict of interest.

References

Byron G (1813) Letter to Annabella Milbanke. Quoted by Woolley B (1999) The bride of science: romance, reason and Byron's daughter. MacMillan, London, p 28

Crick F, Koch C (2005) What is the function of the claustrum? Philos Trans R Soc B 360:1271–1279

Dehaene S (2014) Consciousness and the brain: deciphering how the brain codes our thoughts. Viking Penguin, New York

Dennett D (1991) Consciousness explained. Little Brown, New York

Empson W (1930) Seven types of ambiguity. Chatto and Windus, London, p 9

Fodor J (1998) In: critical condition: polemical essays on cognitive science and the philosophy of mind. MIT Press, Cambridge, p 83

Fodor J (2004) You can't argue with a novel. London Review of Books, London, p 31 (March 3 issue)

Frankish K (2016) Illusionism as a theory of consciousness. J Consciousness Stud 23(11–12):11–39

Gregory R (2011) Gregorian reflections. Image available at http://www.richardgregory.org/gregorian-reflections.htm

Harvard Natural Sciences Demonstrations (2016) Synchronization of metronomes. Available at https://youtu.be/Aaxw4zbULMs

Humphrey N (1992) A history of the mind. Chatto and Windus, London

Humphrey N (2000) One-self: a meditation on the unity of consciousness. Soc Res 67:32–39

Humphrey N (2006) Seeing red: a study in consciousness. Harvard University Press, Cambridge

Humphrey N (2011) Soul dust: the magic of consciousness. Princeton University Press, Princeton

Humphrey N (2016a) A riddle written on the brain. J Conscious Stud 23(7–8): 278–287

Humphrey N (2016b) Redder than red: illusionism or phenomenal surrealism. J Conscious Stud 23(11–12):116–123
Krisztin T (2008) Global dynamics of delay differential equations. Period Math Hung 56:83–95
Maynard J (2015) Is the universe a hologram? Holographic principle suggests 'yes'. Tech Times, New York, 27th April
McGinn C (1993) Consciousness and cosmology: hyperdualism ventilated. In: Davies M, Humphreys, G W (eds) Consciousness. Blackwell, Oxford, p 160
McGinn C (1999) The mysterious flame: conscious minds in a material world. Basic Books, New York
Minsky M (1986) The society of mind. Simon & Schuster, New York
Nagel (2012) Mind and cosmos: why the materialist neo-Darwinian conception of nature is almost certainly false. Oxford University Press, New York. p 35
Newton I (1671) A letter from Mr. Isaac Newton containing his new theory about light and colours. Philos Trans R Soc 6:3075–3087
Orr H (2013) Awaiting a new Darwin. New York Review of Books, New York (Feb 7 issue)
Searle J (1992) The rediscovery of the mind. MIT Press, Cambridge, p 71
Tononi G (2012) The integrated information theory of consciousness: an updated account. Arch Ital Biol 150:56–90