Delusions of consciousness.

Journal of Consciousness Studies, 23, 52-64, 2016
also in *Illusionism*, Ed. Keith Frankish, 23, 52-64, 2017

Susan Blackmore

This is an invited paper for a Special Issue of the *Journal of Consciousness Studies*.

Keith Frankish wrote an introductory article, 'Illusionism as a theory of consciousness', outlining the general view that phenomenal consciousness is an illusion. Nearly twenty responses follow by authors including Dan Dennett, Michael Graziano, Nicholas Humphrey and Jesse Prinz among others.

Here is a version of mine as submitted. It may have been edited before publication. So please do not quote from this version.

Abstract

Frankish's illusionism aims to replace the hard problem with the illusion problem; to explain why phenomenal consciousness *seems* to exist and why the illusion is so powerful. My aim, though broadly illusionist, is to explain why many other false assumptions, or delusions, are so powerful. One reason is a simple mistake in introspection. Asking, 'Am I conscious now?' or 'What is consciousness?' makes us briefly conscious in a new way. The delusion is to conclude that consciousness is always like this instead of asking, 'What is it like when I am not asking what is it like?' Neuroscience and disciplined introspection give the same answer: there are multiple parallel processes with no clear distinction between conscious and unconscious ones. Consciousness is an attribution we make, not a property of only some special events or processes. Notions of the stream, contents, continuity and function of consciousness are all misguided as is the search for the NCCs.

In his clear and helpful survey of the illusionist position, Frankish convincingly argues for taking illusionism seriously and for replacing the hard problem with the illusion problem. This is timely and welcome. From teaching consciousness courses to undergraduates for many years I know how hard it is to get students to see round their strongly held intuitions about dualism, zombies, the knowledge argument and so on. Illusionism cuts through these intuitions and redirects the problem towards asking how and why we have these intuitions and why we seem to be conscious in the way we do. I do not see that any illusionist theory has entirely done away with the hard problem and in the mean time I am being cautious.

The illusionist research programme is clearly worth pursuing – indeed it may be the only research programme worth pursuing. Yet if I understand Frankish correctly, I am not quite an illusionist in his sense. Frankish describes "the basic illusionist claim that introspection delivers a partial, distorted of view of our experiences, misrepresenting complex physical features as simple phenomenal ones." I agree that introspection provides a distorted view but not that it necessarily ends up with 'simple phenomenal ones'. Introspection may sometimes do this and sometimes do the opposite as, for example, in change blindness. This reveals that people believe their own visual worlds contain far more information than they do. This is why I titled the original paper on change blindness, 'The richness of our visual world is an illusion' (Blackmore *et al* 1995). Subsequent studies on change blindness blindness showed that people really do believe they have more information available than they have (Levin *et al* 2000) . In other words they are convinced that they are experiencing a very complex and detailed inner visual world when in fact they are relying on their ability to look again (Simons and Rensink 2005).

For this and other reasons I refer to consciousness as illusory and have tried to work out how and why the illusions arise. Yet my enterprise is far less ambitious than Frankish's. I am less

concerned with explaining 'why experiences seem to have phenomenal properties' (his 'illusion problem') and more concerned with why they lead to false theorising. Many researchers seem, despite their fervent denials, to be Cartesian materialists (Dennett 1991) and mired in lurking dualist assumptions. This hampers research by leading them to ask the wrong questions and look in the wrong places for answers. So my aim is not the bold one of escaping the hard problem but the more limited one of exposing and understanding the power of these false ideas. Because this theorising is largely intellectual the assumptions might be called 'delusions' rather than 'illusions'.

Delusionism

Sitting here at my desk I can feel the keys beneath my cold fingers, hear birds singing outside the window and see a wealth of colour in the trees and flowers and sky. When introspecting on this it is all too easy to leap from immediate sensations to theorising about what 'must' be going on. Failing to stop at sensations of cold or dappled green or bird song, we readily assume a 'me' who is the subject of a stream of conscious experiences, and that these arise from (or emerge from, are produced by, or even 'are') a few brain processes while everything else remains 'unconscious'. Yet all this is fantasy. It is these proto-theories, not the cold, the green or the singing, that constitute the delusion.

Frankish asks why the illusion of phenomenality is so powerful. I would ask the related, but different, question of why this deluded theorising is so tempting and so powerful. The answer, I suggest, is amusingly simple, if counter-intuitive. Ask yourself this question:

'Am I conscious now?'

I guess that your answer is 'yes'. It is very hard, though not impossible, to answer truthfully 'no' (more of this later). So, whenever we ask this question, we reply, 'Yes – I am conscious now'. We can ask many times and always get the same answer, making it easy to conclude that life is always like this, that I am continuously present and experiencing a stream of thoughts and perceptions. So we assume a continuity and unity which is simply not true. The illusion is powerful because it is so hard to answer a different question – what is it like the rest of the time? What is it like when I am not asking what is it like?

This illusion is similar in structure to why 'the richness of our visual world is an illusion'. The reason we seem to see a richly detailed and picture-like view of the world is that wherever we look we see rich detail. We can always look again and the detail appears 'just in time', so it *seems* that the whole picture is in our minds rather than where it always was, out in the world (O'Regan and Noë 2001). This illusion is powerful because it is hard to answer the question, 'What does something look like when I am not looking at it?'

There is thus something very curious about the nature of consciousness – that looking into consciousness reveals only what it is like when we are looking into it – and most of the time we are not. So introspecting on our own minds is thwarted by the very fact of introspecting.

William James had a wonderful metaphor for this more than a century ago. Trying to observe the "flights" as well as the "perchings" in his "stream of consciousness", he said "The attempt at introspective analysis in these cases is in fact like seizing a spinning top to catch its motion, or trying to turn up the gas quickly enough to see how the darkness looks." (James 1890, 244).

The modern equivalent might be trying to open the fridge door quickly enough (Blackmore 2012, O'Regan 2011). Similarly, asking "Am I conscious now?" or "What am I conscious of now?" or even "What is phenomenality?' or, 'What is consciousness?" can feel like turning on a light, but is that light always on? And if not then what is the darkness like?

With a fridge we can find out, for example by drilling a hole in the side and looking in when the door is still closed, or by understanding how the switch works. To some extent we can do this with brains too. We can look inside with electrodes or scanners and see what is going on.

We do not find a light and a switch or a cold dark cupboard full of food. But nor do we find a conscious self, a place where a conscious self could be, or any roles for it to play. We find billions

of neurons connected up in trillions of ways with vast numbers of parallel processes going on simultaneously. There is no central processor, no place for an inner observer commanding the action and, and no show in the Cartesian Theatre (Dennett 1991). To all appearances there are just lots and lots of neurons firing and chemicals moving about. If we ask which are the conscious ones how can we tell?

Can we take a different tack and look into the darkness personally, by training our introspection to look more carefully? I have described in detail my many attempts to do this using meditation combined either with formal Zen koan practice or with persistently asking questions of my own devising (Blackmore 2011). Hundreds of students on my consciousness courses have explored these questions too (Blackmore 2010/11). Once they get used to the sensation of becoming more conscious, or even 'waking up', by asking, 'Am I conscious now?', a second question naturally pops up: 'Was I conscious a moment ago?' or 'What was I conscious of a moment ago?'. Asking these, I suggest, leads to a curious discovery – that we ourselves do not know the answer.

A striking example of this occurred just yesterday. I was climbing a steep hill with rough and unequal steps, noticing the changing rhythm as I climbed; one left, two right, two left, one right. I asked myself (as I so often do), 'Am I conscious now?' (Yes, I'm conscious of the climbing) and then, 'What was I conscious of a moment ago?'.

Obviously there was the rhythm of climbing, but once I had asked the second question I could also remember hearing – or having had the feeling that someone or something had been hearing – the scrunching of my feet on the rough surface which I had not noticed until I asked. Then I remembered hearing my own laboured breathing going in and out, the effort in my legs, and the burning sensation of the sun on my shoulders. In an odd way these now seemed to have been as conscious as the rhythm, although in another way they seemed not to have been conscious at all because I had only just noticed them. I was musing briefly on this familiar and long-practiced oddity when suddenly, in a startled flash, I remembered something else.

Just two steps before I had lifted my arm, looked at my watch, seen that it was just before noon, decided that I was on track for where I was going, and dropped my arm again. The memory was clear, detailed and vivid. So was I conscious of checking the time?

I might answer yes, because now I could clearly remember doing it, even to the look of my arm in the sun and the position of the hands on my watch. I might answer no, because the whole memory seemed to come 'into my consciousness' afterwards only because I asked the second question. Which is right? I do not know, and if I do not know then how can the answer be discovered? Indeed, is there an answer at all?

I suggest not. To explore further, let me ask some other, related, questions. Did I consciously decide to check the time? I had no memory of intending or planning the action but I might just have forgotten the intention. How can I find out?

Was performing the actions conscious at the time? The same applies.

Here's a harder, but important, one. Were the brain processes underlying the actions conscious brain processes? Clearly the actions are quite complex and involved multiple perceptual, cognitive and motor processes, just as noticing the rhythm of climbing the steps does. Can we find out which of these were conscious or unconscious processes?

We could try.

We might look to the concept of attention. If attention is thought of as resource allocation then considerable attention must have been paid to looking at my watch and registering the time, and presumably it would be possible, at least in principle, to measure the amount of work going into all this. So, although I seemed to be conscious only of the rhythm of climbing, resources were clearly split, as they are in the familiar 'unconscious driving phenomenon', and it would be hard to say that far more resources were being used for one than the other. This means that we cannot use the amount of attention in any simple way to decide which processes were conscious and which unconscious.

We might look to the popular global workspace theories for an answer. According to Baars' (1988) original formulation of GWT the contents of consciousness are the processes currently in the GW, or 'on the stage', and they become conscious by virtue of being globally broadcast to the rest of the unconscious brain. According to the later 'neuronal GWT', information becomes conscious when the long-distance connectivity of 'workplace neurons' makes it widely available and this is what we experience as a conscious state (Dehaene and Naccache 2001). But what does this mean? And how and why does this broadcast make information conscious?

There are two radically different ways of interpreting GWTs. The first, and more common, is that when the contents on the stage are broadcast they 'become conscious'. This version of Cartesian materialism (Dennett 1991) retains the hard problem. Something magical has happened to the previously unconscious contents so that they are now conscious ones. They have gained, or have become, qualia or phenomenality. The alternative is that nothing more happens to them at all. Being broadcast is all there is. This is what Dennett (2001) means by 'cerebral celebrity' or 'fame in the brain'. Just as there is nothing more to being famous than being well known by lots of people, so there is nothing more to consciousness than being widely available in the brain. This availability has consequences for later actions and perceptions, including the ability to talk about what happened, to attribute consciousness to the sensations and to base further actions upon them.

Applying GWT to my example, we might say that before I asked the question the only processes on the stage or in the GW were those involved in climbing and counting the steps; these were being broadcast to the rest of the unconscious brain while those involved in checking the time were not.

Now we can see the problem. Had I not asked the question I would have totally forgotten the way my arm looked as it rose and fell. Yet this action would certainly have had consequences for later brain processes, thoughts and actions. If, when I arrived at the top of the hill, I had wondered what time it was I would have known that it was just gone twelve. So what are we to say? That the broadcast from looking at my watch was, for some reason, not sufficient or not of the right sort of broadcast, to count as being 'on the stage' or 'in consciousness'? If so, for what reason?

I suggest that the whole idea of the GW, and its popularity, arises from the illusion I have described. Whenever we ask about consciousness, a temporary unity of a set of thoughts and perceptions is constructed and is linked to a representation of self as a continuing observer (Metzinger 2009). This we call the contents of our consciousness while everything else is called 'unconscious'. GWT nicely captures this intuition, which is based on a momentary and misleading situation.

Returning to my example, as I asked the first question a self-model was briefly constructed of me climbing, counting and looking at the ground but not including looking at my watch. If we could look inside the brain in sufficient detail I guess we would see some of the hill-climbing processes linked to the body schema, and to self-modelling and questioning processes while the watch-looking and many other processes were going on separately. When I asked the second question, lots more processes, including the watch-looking, were combined to make an even more complex whole. When I stopped asking about consciousness and got on with climbing the hill the temporary coherence dissolved and normality resumed. The multiple parallel processes just carried on, none linked to a model of self as observer; none either in or out of consciousness; none either conscious or unconscious.

I conclude that there is no intrinsic difference between conscious and unconscious processes, nor between conscious and unconscious actions or perceptions. Rather, consciousness is a fleeting attribution that we make if and when we ask about it, either when asking such questions as, 'What am I conscious of now?' or in retrospect when we think about the past. This implies that most theories of consciousness address only rare moments in our lives.

So what was it like before we asked? We might try to find out using either the methods of neuroscience or of disciplined introspection. Amazingly enough, both come to the same answer:

that there are lots and lots of parallel processes going on and no obvious way to tell which were conscious.

Is it possible to find out? Are there actual, but unobservable, facts about which processes, thoughts or actions were conscious at any time? I say no. The neural correlates of all these would in principle be observable but with no way of distinguishing between conscious and unconscious ones either by objective measures or in subjective experience. The distinction is meaningless because consciousness is an attribution we make, not a property of events, thoughts, brain processes or anything else.

If this is so, we must reject not only many folk-psychological beliefs but also many common phrases used in the literature, with interesting implications for the science of consciousness.

Implications for the study of consciousness

There is no stream of consciousness

William James coined phrase 'the stream of consciousness' in his classic 1890 work *The Principles of Psychology* and it has been used ever since to imply something like the folk psychological idea that we are conscious subjects experiencing an ever-changing but unified flow of ideas, thoughts, perceptions and intentions. I suggest that consciousness appears as a stream only when we reflect on it as such. The rest of the time, multiple parallel processes carry on, sometimes interacting with each other, often not. When we ask 'What am I conscious of now?' or 'What is it like being me now?' some are gathered together and the answer appears stream-like. There are memories of recent perceptions and thoughts and, if we remain mindful for a few moments, a changing array of new perceptions and thoughts coming along. There is a powerful sense of someone who is experiencing this stream. The illusion is to believe there is also a stream like this when we are not inquiring (Blackmore 2002).

There are no contents of consciousness.

The idea that consciousness has contents is perfectly natural and easy to imagine. It appears in many guises; as the contents of the stream, the items on the stage of the GW, the features illuminated by the spotlight of attention, and the show in Dennett's mythical Cartesian Theatre.

There is a new problem revealed here. I have used the words 'item', 'feature' and 'show' but what is really being referred to? When talking in brain terms people often say 'processes', 'representations', 'neural patterns' or just 'information': in psychological terms, 'items', 'ideas', 'perceptions', 'thoughts' or even just 'things'. I find myself reverting to 'things' when struggling to understand what is supposed to be in the stream, in the GW or on the stage.

Ignoring this problem, these 'things' must start by being unconscious, perhaps building up ready to enter the GW or moving into the spotlight of attention. Then they become conscious before finally leaving and becoming unconscious again, perhaps disintegrating as they do so. What could it mean for these things (items, processes, thoughts, information or whatever) to become conscious? Do they change from being objective things to subjective experiences? Do they suddenly become phenomenal or qualia laden (Ramachandran and Hirstein 1997) or get qualia attached (Gray 2004)? The hard problem has not disappeared but has merely been restricted to 'contents' inside the hypothetical stream, GW, or stage. The impression of a space that contains everything we are conscious of at any time seems natural and obvious, but consciousness is not a container (Blackmore 2002).

The unity of consciousness

This phrase is often taken to mean that at any time all the things I am conscious of form some kind of collective. It can certainly seem that way, as if the sensations of sitting here, the pictures on the wall in front of me, the sounds around me, all form some kind of unified field. We might guess that in neural terms (this is, of course, an empirical question), attention brings together some of the ongoing parallel processes and so creates temporary coalitions that provide a sense of unity. Yet it takes only a little careful observation to see that as my attention shifts these sensations come and

go and that these acts of attention pull some together and let others disappear. This is the only sense in which there is 'unity of consciousness'.

The continuity of consciousness

This is the easiest idea to demolish by trying to 'look into the darkness'. Every time I 'turn on the light' by asking myself what my consciousness is like, it seems to be a flowing stream of ever-changing, unified contents, much as it was last time I looked. That's fine. That's how it seems, and how it seems is what we are trying to explain.

The illusion is to leap from that repeated observation to the conclusion that consciousness is always that way; that the stream of contents continues without break when it does not.

The neural correlates of consciousness

The hunt for the NCCs is probably the most popular research paradigm in consciousness studies (Metzinger 2000). From experiments in binocular rivalry to Baars' (1988) 'contrastive analysis' the idea is to find the difference between those neural processes that are conscious and those that are not. If my analysis is correct it is obvious that this whole approach is doomed.

I do not doubt that neuroscientists can find, in ever greater detail, the NCs of specific actions, thoughts, perceptions, and so on. They can also look for the NCs of asking such questions as, 'What is it like to be me now?' or 'What is consciousness?' and of making attributions of consciousness to past or present thoughts and perceptions. But they will never find the NCs of an extra added ingredient – 'consciousness itself' – for there is no such thing.

The function of consciousness

We humans are conscious, so the argument goes, therefore consciousness must have evolved for a reason and must have a function. Many functions have been proposed, including error monitoring (Crook 1980), an inner eye (Humphrey 1986), saving us from danger (Baars 1997), late error detection (Gray 2004), and giving a point to survival (Velmans 2009). Yet this type of argument can be like believing in the possibility (not the conceivability) of philosophical zombies. It assumes some special extra thing called consciousness 'itself' in addition to all the underlying functions and processes. It assumes that there might have been creatures with and without this special extra and that the ones who had it survived, reproduced and passed on their genes more effectively than those without.

If my analysis is right then this makes no sense. Consciousness, as subjectivity, is not a force that can do anything or have effects or consequences that could form the basis of any evolutionary function. It did not evolve. What evolved was an intelligent creature with the capacity for selective attention, language and introspection, the ability to call some of its own actions, thoughts or perceptions conscious, and hence to fall for the illusion that consciousness has a function.

As Frankish points out, this offers a new perspective on the function of consciousness, because we can ask what function such illusions serve or how they might be adaptive. In his later work, Humphrey (2006, 2011) spells out one answer; that from monitoring their own responses to the world creatures internalised sensations, leading to an inner 'magical-mystery show' and an illusory soul. Consciousness seems so important to us because '*it is its function to matter*' and to seem other-worldly and mysterious (2006 p131).

This is a radically different approach. Yet, like these other theories, it still assumes that the ultimate function is biological survival; that consciousness, or the illusions of consciousness, must benefit human genes. But this is not the only possibility.

Elsewhere I have argued that the benefit accrues more to memes than genes (Blackmore 1999). To survive and reproduce memes need first to find homes in human brains and then find ways to get passed on. A meme that becomes 'my' idea, preference, need, favourite joke or special song has an evolutionary advantage over one that does not. Memes that make up the stories I tell about myself or the opinions I express have an advantage, and they build up into what I have called the 'selfplex', a co-adapted complex of memes that all thrive better together than apart. A self with

strong opinions, lots of ideas and a need for status and power makes an effective meme spreader even if it is not the continuous, unified and powerful subject of experience it models itself as being.

Is the illusory self, as Dennett (1991) would have it, a 'benign user illusion'? I think not. It is arguably a 'malign user illusion' and the source of much suffering and misery (Blackmore 2010/11). This is the self who craves love, friendship, status, possessions and power. This is the self who gets disappointed, hurt, lonely, angry and resentful. This is the self who wants happiness but when happy fears losing it. This is the self who makes constant comparisons with others and fears other peoples' judgements. It is strange that an illusion can entail so much suffering.

Is it possible to throw off these illusions completely? Mystics and contemplatives for thousands of years have claimed that it is; that greed, hatred and delusion can give way to kindness and equanimity, even if the endeavour takes decades of meditation or years of solitude in a remote mountain cave. Others have discovered similar insights through spontaneous mystical experiences or the use of shamanic brews and psychedelic drugs. Many long-term meditators describe experiences without an experiencer, or nondual states in which experience and experiencer are one. In these states we can ask 'Am I conscious now?' and truthfully answer 'no' because there is no constructed self to be the experiencer (Blackmore 2011). The experienced duality of self and other then disappears along with the hard problem and yet it is hard to make rational sense of this directly experienced nonduality.

Conclusion

As Frankish suggests, the question is not whether illusionism is intuitively plausible, but whether it is rationally compelling. His illusionism aims to explain why phenomenal consciousness *seems* to exist. My own aim is the slightly different one of using introspection to explain why some other delusions about consciousness are so compelling, and perhaps this way to make delusionism more intuitively plausible. Mysteries remain and, as Dennett puts it, a mystery is something we don't yet know how to think about. All I have tried to do is to clear away some of the intuitively appealing but wrong ways of thinking about consciousness in the hope that they may be replaced by better ones.

References

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

Baars,B.J. (1997) *In the Theatre of Consciousness: The Workspace of the Mind.* New York, Oxford University Press

Blackmore, S. (1999) *The Meme Machine,* Oxford, Oxford University Press

Blackmore,S.J. (2002) There is no stream of consciousness. *Journal of Consciousness Studies,* **9,** 17-28

Blackmore, S. (2010) *Consciousness: An Introduction,* Second Edition, London, Hodder Education, and 2011 New York, Oxford University Press,

Blackmore, S. (2011) *Zen and the Art of Consciousness,* Oxford, Oneworld Publications

Blackmore, S. (2012) Turning on the light to see how the darkness looks. In *Consciousness: Its Nature and Functions*, Ed Shulamith Kreitler and Oded Maimon, NY, Nova pp 1-22

Blackmore,S.J., Brelstaff,G., Nelson, K. and Troscianko,T. (1995) Is the richness of our visual world an illusion? Transsaccadic memory for complex scenes. *Perception,* **24**, 1075-1081c

Crook,J. (1980) *The Evolution of Human Consciousness* Oxford University Press

Dehaene S, Naccache L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition. 79, 1–37

Dennett, D.C. (1991) *Consciousness Explained.* London, Little, Brown & Co.

Dennett, D. (2001). Are we explaining consciousness yet?. *Cognition*, *79*(1), 221-237.

Gray, J. (2004) *Consciousness: Creeping up on the Hard Problem* Oxford, Oxford University Press

Humphrey, N. (1986) The Inner Eye. London, Faber & Faber

Humphrey, N. (2006) Seeing Red Cambridge, MA., Harvard University Press

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

James,W. (1890) *The Principles of Psychology,* London; MacMillan

Levin, D. T., Momen, N., Drivdahl IV, S. B., & Simons, D. J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition*, *7*(1-3), 397-412.

Metzinger, T. (Ed) (2000) *Neural Correlates of Consciousness*, Cambridge, MA., MIT Press

Metzinger, T. (2009) *The Ego Tunnel**:** The Science of the Mind and the Myth of the Self,* London, Basic Books

O'Regan, J.K. (2011) *Why Red Doesn't Sound Like a Bell*, New York, Oxford University Press

O'Regan, J.K. and Noë, A. (2001) A sensorimotor account of vision and visual consciousness.*Behavioral and Brain Sciences*, **24**(5), 939-1011

Ramachandran,V.S. and Hirstein,W. (1997) Three laws of qualia: What neurology tells us about the biological functions of consciousness. *JCS,* **4,** 429-457

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in cognitive sciences*, *9*(1), 16-20.

Velmans,M. (2009) Understanding Consciousness. London, Routledge