Daniel C. Dennett

Illusionism as the Obvious Default Theory of Consciousness

Abstract: *Using a parallel with stage magic, it is argued that far from being seen as an extreme alternative, illusionism as articulated by Frankish should be considered the front runner, a conservative theory to be developed in detail, and abandoned only if it demonstrably fails to account for phenomena, not prematurely dismissed as 'counter-intuitive'. We should explore the mundane possibilities thoroughly before investing in any magical hypotheses.*

Keith Frankish's superb survey of the varieties of illusionism provokes in me some reflections on what philosophy can offer to (cognitive) science, and why it so often manages instead to ensnarl itself in an internecine battle over details in ill-motivated 'theories' that, even if true, would be trivial and provide no substantial enlightenment on any topic, and no help at all to the baffled scientist. No wonder so many scientists blithely ignore the philosophers' tussles — while marching overconfidently into one abyss or another.

The key for me lies in the everyday, non-philosophical meaning of the word *illusionist*. An illusionist is an expert in sleight-of-hand and the other devious methods of stage magic. We philosophical illusionists are also illusionists in the everyday sense — or should be. That is, our *burden* is to figure out and explain *how the 'magic' is done.* As Frankish says:

> Illusionism replaces the hard problem with the illusion problem — the problem of explaining how the illusion of phenomenality arises and

Correspondence:
Daniel C. Dennett, Tufts University, Medford, MA, USA.
Email: daniel.dennett@tufts.edu

Journal of Consciousness Studies, **23**, No. 11–12, 2016, pp. 65–72

> why it is so powerful. This problem is not easy but not impossibly hard either. The method is to form hypotheses about the underlying cognitive mechanisms and their bases in neurophysiology and neuroanatomy, drawing on evidence from across the cognitive sciences. (p. 37)

In other words, you can't be a satisfied, successful illusionist until you have provided the details of how the brain manages to create the illusion of phenomenality, and that is a daunting task largely in the future. As philosophers, our one contribution at this point can only be schematic: to help the scientists avoid asking the wrong questions, and sketching the *possible* alternatives, given what we now know, and motivating them — as best we can. That is just what Frankish has done.

He distinguishes and assesses the different versions of illusionism, illuminating the paths that led to them. His account of my own view is flawless, and I was enlightened by many of his remarks about the ideas of other illusionists I thought I fully understood. He is particularly good on the various motivations of the variations, but I want to stress the motivation for the overall strategy of taking illusionism — in any form — seriously. And here, once again, it is useful to take a glance at stage magic. In today's world, if not in the Dark Ages or even the Renaissance, the standard, default assumption about any feat of stage magic we encounter is that it is (somehow) the product of everyday physical causes, involving perhaps magnets and electricity, or even holograms, but not psychokinesis, clairvoyance, or the assistance of any poltergeists, goblins, or other supernatural agents. In other words, it is *stage* magic, not 'real magic'. (In his excellent book on Indian street magic, *Net of Magic: Wonders and Deceptions in India*, Lee Siegel writes,

> 'I'm writing a book on magic,' I explain, and I'm asked, 'Real magic?' By *real magic* people mean miracles, thaumaturgical acts, and supernatural powers. 'No,' I answer: 'Conjuring tricks, not real magic.' *Real magic*, in other words, refers to the magic that is not real, while the magic that is real, that can actually be done, is *not real magic*. (Siegel, 1991, p. 425))

The common understanding is that magicians are doing tricks, not engaging in sorcery, and moreover, the default — but defeasible — assumption about any feat of stage magic is that it doesn't avail itself of quantum entanglement (Einstein's 'spooky action at a distance') or the creation of any new subatomic particles or other forces that would institute a revolution in physics. Hey, it's just stage magic!

In short, when it comes to stage magic we assume, until positively shown otherwise, that the effects are achieved by some hard-to-imagine concoction of everyday physical causes and effects. Here is where anybody, philosopher or scientist or visionary, is apt to suffer a failure of imagination and mistake it for an insight into necessity. As the noted illusionist Jamy Ian Swiss has said, 'No one would ever think that we would ever work this hard to fool you. That's a secret, and a method of magic' (2007, the e.g.conference, www.egconf.com/videos/how-magic-works). This is not just an interesting observation. It draws attention to a fact that puts *all* philosophers on notice: nobody would, or should, take seriously a would-be explainer of stage magic who declared that it was just undeniably, intuitively obvious that no *possible sequence of ordinary physical events* could account for the feat just observed. We philosophical illusionists say that before you run off half cocked with theories about consciousness as one sort or another of 'real magic', you should try to explain it all as an illusion engendered by nature. When philosophers rely on what follows from their most undeniable *intuitions* about consciousness, before exhausting the physical possibilities, they are not adopting a sound method of enquiry but simply indulging themselves. Imagine Chalmers' declaration that phenomenal consciousness is a datum, transposed into the claim that a lady-sawn-in-half is a datum, or the claim that we are directly acquainted with the real presence of a lady-sawn-in-half. You may *think* you're directly acquainted with this, but that's a fact of personal psychology, at best an unshakable intuition, not a datum. You could be wrong, and until we have canvassed the alternatives, we should put your intuition on the back burner, not honour it. Or consider Searle's italicized dictum that '*where consciousness is concerned the existence of the appearance is the reality*' (quoted by Frankish, this issue, p. 32). Maybe, and maybe not. Searle apparently thinks that this is crushingly obvious, and he is not alone. When we know more about the brain's activities we *will see* if we can eliminate the prospect of the brain creating an illusion of 'appearance', of phenomenality. You can't just declare, as a first principle, that this is impossible.

Illusionism may today seem 'incredible, even ludicrous' (Frankish, p. 37), but if and when it is eventually fleshed out with details, those gullible folks who fell for the trick will discover that they have industriously concocted their card castle 'theories' for nothing. I put 'theories' in scare quotes because most philosophical theories are just definitions defended, with no aspiration to make novel predictions but

rather just to assign the phenomena covered by the 'theory' to some category or other. They at best clarify and articulate the implications of the everyday concepts involved. A weakness of such 'theories' is that, since they are largely driven by shared folk intuitions, they are always playing catch-up, seeing if they can accommodate newly discovered but unanticipated scientific discoveries, instead of pioneering perspectives from which new empirical questions can be asked and answered. When counter-intuitivity counts against an hypothesis, with nothing but consistency to other intuitions to support it, there is little chance of making progress. Illusionism boldly discards a host of all too comfortable intuitions. As Frankish says, 'The question is not whether illusionism is intuitively plausible, but whether it is rationally compelling' (p. 37).

Illusionism, I am saying, should not be seen as a lame attempt to deny the obvious, but as the leading contender, the default view that should be assumed true until proven otherwise. (I grant that my whimsical title, 'Quining Qualia', lent unintended support to the perception that illusionism is a desperate and incredible dodge, and for that little joke I now repent.)

From this perspective, we can see that philosophers of mind who are *not* illusionists are prematurely encouraging scientists to worry about the wrong questions, artefactual problems like those that would arise for any scientists trying to uncover the details of the quantum-entanglement theory of teleporting the beautiful assistant from one trunk to another or trying to reconcile the actual presence of the ten of diamonds in their pocket where they put it with its manifest presence on the table. It is a remote possibility that we will have to fall back on quantum physics or multiple universes to account for some mind-boggling bit of magic, but first, let's try the conservative route.

One of the difficulties illusionists must tackle is how to temporize with terminology until the facts are in, and here the big weak spot, in my opinion, is the term 'representation'. A fanatic about proper procedure would insist that one *never* use the term without saying, very clearly and in detail, who or what the user/audience is for the so-called representation. But we simply cannot meet that burden *in detail* yet. So we are tempted to leave the 'end user' slot unspecified, which amounts, in theories of consciousness, to evading what I call the Hard Question: 'And then what happens?' (Dennett, 1991, p. 255). It makes a big difference whether a postulated representation is *for* modulating hand–eye coordination or *for* identification or categorization of objects, or *for* … conscious experience. Today we — most of us —

are comfortable with systems of *unconscious* representations that influence, specify content, orient, direct memory retrieval, etc. That is as good as gospel in cognitive science. These are representations *in us* that contribute to our cognitive talents without being *for us.* (In this regard they are no different from the representations of blood sugar level or vitamin deficiency that modulate our digestive systems without engaging cerebral cortex at all.) But at some point, as Frankish puts it, we must describe

> the sensory states that are the basis for the illusion. On most accounts, I will assume, these will be representational states, probably modality-specific analogue representations encoding features of the stimulus, such as position in an abstract quality space, egocentric location, and intensity. (p. 19)

Filling in these details will require answering a host of questions that Frankish raises:

> Is introspection sensitive only to the content of sensory states, *or* are we also aware of properties of their neural vehicles? Do the reactions and associations evoked by our sensory states *also* contribute to the illusion of phenomenality?… Are sensory states continually monitored *or* merely available to monitoring? Is the introspectability of sensory states a matter of internal access and influence *rather* than internal monitoring? (p. 19, my italics)

I submit that, when we take on the task of answering the Hard Question, specifying the uses to which the so-called representations are put, and explaining how these are implemented neurally, some of the clear alternatives imagined or presupposed by these questions will subtly merge into continua of sorts; it will prove not to be the case that content (however defined) is sharply distinguishable from other properties, in particular from the properties that modulate the 'reactions and associations evoked'. The distinction between continual monitoring and availability to monitoring will also vanish (it already has vanished in many computer programs). Suppose that neural signals with content (*consistent with*) *dog here now* arrive at location X in the cortex every 20 milliseconds, thereby preventing location X from *enquiring* whether there seems to be a dog here now. Is location X continually monitoring for dog-presence, or just lulled into a complacent state of disinterest? (*Cf.* in what sense is 'the background' *there* in unattended 'phenomenal consciousness'?) This may not be a well-motivated question when we learn more about the mechanisms involved (see Cohen, Dennett and Kanwisher, 2016). Frankish's

questions are good questions, but that doesn't mean that they will all have crisp answers. The answer may well be that these distinctions do not travel well when we abandon what Sellars (1962) calls the manifest image and get down in the trenches of the scientific image.

Here is a sentence that tantalizes me, in Daniel Wegner's book, *The Illusion of Conscious Will*:

> We can't possibly know (let alone keep track of) the tremendous number of mechanical influences on our behavior because we inhabit an extraordinarily complicated machine. (Wegner, 2002, p. 27)

This frankly Cartesian formulation exposes the allure of the traditional manifest image: there is a place in the brain where 'our minds portray their operations to us, then, not their actual operation' (*ibid.*, p. 96). Wegner was silent about how this portrayal (to 'us') is accomplished, so it is not clear if he counts as an illusionist. I think he should count, since he gave a good if partial answer to the Hard Question while eschewing all questions about the 'phenomenality' of the portrayals. As he recognized, the effects he provoked in his experiments could be accounted for as stage magic.

The seeds of illusionism can already be discerned in U.T. Place's pioneering article, 'Is Consciousness a Brain Process?' (1956). Place was so bold as to identify the *denial* of illusionism as a fallacy, the *phenomenological fallacy*:

> [T]he mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field'. (*ibid.*)

How then does one avoid the phenomenological fallacy? J.J.C. Smart offered an answer in 1959 and elaborated on his answer in 1963:

> The man who reports a yellowish-orange after-image does so in effect as follows: '*What is going on in me is like what is going on in me when* my eyes are open, the lighting is normal, etc., etc., and there really is a yellowish-orange patch on the wall.' (Smart, 1963, p. 94)

As Smart pointed out way back then, it is quite possible for a mechanism to be able to discern or discriminate when something going on in it is like something else without having any idea just wherein that similarity resides. If we fill the heads of people with such mechanisms, suitably organized and orchestrated, they can provide a large

part of the answer to 'And then what happens?' without ever postulating anything like phenomenality.

How might this go? When you seem to see a red horizontal stripe (as a complementary-colour after-image of a black, green, and yellow American flag), there is no red stripe in the world, no red stripe on your retina or in your brain. There is no red stripe anywhere. There is a 'representation' of a red stripe in your cortex and this cortical state is the source, the cause, of your heartfelt conviction that you are in the presence of a red stripe. You have no privileged access to how this causation works. We have a good theory of how colour perception works, with its opponent processes and refractory periods, so you can probably explain the early or distal links in the causal chain from eyeball to conviction, but you simply don't know what the proximal or immediate causes are that put you into a state of subjective conviction and the attendant further sequelae ('and then what happens?'). (And this is true of your access to normal, not illusory, vision as well, of course.)

The red stripe you seem to see is not the *cause* or *source* of your convictions but the *intentional object* of your convictions. In normal perception and belief, the intentional objects of our beliefs are none other than the distal causes of them. I believe I am holding a blue coffee mug, and am caused to believe in the existence of that mug by the mug itself. The whole point of perception and belief fixation is to accomplish this tight coalescence of causes and intentional objects. But sometimes things go awry. Suppose a gang of hoaxers manage to convince you, by a series of close encounters, that there is a space alien named Zom who visits you briefly, speaks to you on the phone, etc. The *causes* of your various Zom experiences can be as varied as can be, so that nothing at all in the world deserves to be identified as Zom, the intentional object of your beliefs.

What are intentional objects 'made of'? They're not made of anything. When their causes don't coalesce with them, they are fictions of a sort, or illusions. We don't postulate *fictoplasm* as the substance from which Sherlock Holmes and Hamlet are composed, and likewise we shouldn't postulate *figment* as the stuff of 'phenomenal properties' (Dennett, 1991). When there is a red stripe in the world, the redness is a complex physical property of the stripe; when there just seems to be a red stripe in the world, that very same property is *represented* as being present by some team of brain agents that are the cause of your false conviction. The eternally tempting mistake is to fall for this chain of inference:

> It seems to me as if there's a red stripe projected out onto that wall, but there is no such red stripe out there,… *so it must be in here.*
>
> And even when there is a red stripe out there in the world that I see, my seeing it must involve an intermediary 'phenomenal' red stripe in my consciousness.

We illusionists advise would-be consciousness theorists not to be so confident that they *couldn't* be caused to have the beliefs they find arising in them by mere neural representations lacking all 'phenomenal' properties. Of course they could; just ask stage magicians — illusionists in the everyday sense — who specialize in provoking false but passionately held beliefs in things that they seemed to see but didn't see.

References

Cohen, M., Dennett, D.C. & Kanwisher, N. (2016) What is the bandwidth of perception?, *Trends in Cognitive Sciences*, **20** (5), pp. 324–334.

Dennett, D.C. (1991) *Consciousness Explained*, Boston, MA: Little, Brown.

Place, U.T. (1956) Is consciousness a brain process?, *British Journal of Psychology*, **47**, pp. 44–50.

Sellars, W. (1962) Philosophy and the scientific image of man, in Sellars, W., *Science, Perception and Reality*, London: Routledge & Kegan Paul.

Siegel, L. (1991) *Net of Magic: Wonders and Deceptions in India*, Chicago, IL: University of Chicago Press.

Smart, J.J.C. (1959) Sensations and brain processes, *The Philosophical Review*, **68** (2), pp. 141–156.

Smart, J.J.C. (1963) *Philosophy and Scientific Realism*, London: Routledge & Kegan Paul.

Swiss, J.I. (2007) A lecture at the e.g.conference, [Online], www.egconf.com/videos/how-magic-works.

Wegner, D.M. (2002) *The Illusion of Conscious Will*, Cambridge, MA: MIT Press.