This is the author's eprint of a chapter published in C. Sosis (ed.), *Too Weird to Believe, Too Plausible to Deny* (pp. 214-223), Wiley-Blackwell, 2025. It may differ in minor ways from the print version. The definitive text is available at https://www.taylorfrancis.com/chapters/edit/10.4324/9781003478300-25/ Version 1, 3/4/26

Consciousness is not what it seems

Keith Frankish

Seeming and Reality

Is there anything we can be absolutely certain of? When the seventeenth-century philosopher René Descartes asked himself this, he famously answered that there was only one thing that was certain: the contents of his own mind. He could doubt that there was an external world (perhaps he was having a vivid dream) and even that he had a body, but he couldn't doubt the reality of the thoughts and experiences occurring in his mind. He argued that this indubitable fact formed a basis on which all knowledge could be built (Descartes, 1641/1984).

This is a tempting idea. You can't be sure what the world is like, but you can be sure what your *experience* of it is like — your consciousness. For example, we often make mistakes about the colors of objects. (Indeed, in a sense, objects themselves don't have colors, just dispositions to reflect light in certain ways.) But we don't make mistakes about our color *experience*. Things may not be the way they seem, but the way they seem is the way they seem. When it comes to consciousness, there is, as some philosophers say, no *appearance/reality gap*. Let's call this the *Cartesian view* of consciousness (after Descartes). Many people think the view is obviously right, but I'm going to argue that it's seriously flawed.

One thing is clear. We take ourselves to have an *authority* about our own experiences that we don't have about other things. We readily make claims about the world around us. We say, for example, that a dress is yellow, that our food is tasteless, that the train we are on is moving, and so on. But we know that these claims may be wrong. Indoor lighting can affect our color judgements, a head cold can distort our taste judgements, the motion of things around us can trick us into thinking we are moving, and so on. If we suspect that this is happening or just want to be cautious (as Descartes did), we modify our claims by changing "is" to "seems." We say that the dress *seems* yellow, that our food *seems* tasteless, that the train *seems* to be moving, and so on. And we take these claims to be secure. Someone can convince me that the dress isn't really yellow, but they can't convince me that it doesn't *seem* yellow to me. If I sincerely report that it seems yellow to me, then that's the end of it. Sincere, careful reports of how things seem are authoritative. The question is *why* they have this status.

The Theater of the Mind

The Cartesian view offers one answer. It says that reports of how things seem are authoritative because they express knowledge of our conscious experiences, and this

knowledge is certain. Conscious experience (the story goes) involves awareness of private mental qualities — mental versions of the colors, sounds, smells, tastes, and other qualities we attribute to external things. These qualities, which philosophers call *qualia* (singular *quale*), constitute the way things seem to us, and, provided we attend carefully, we can make no mistake about them. Thus, when we report that something seems yellow to us, our claim is authoritative because we are directly acquainted with something — a quale — that *really is* yellow.

On this view, then, consciousness is a sort of mental interface, where information gathered by our senses is assembled and presented in a multi-sensory qualia show. The philosopher Daniel Dennett calls this supposed qualia show the *Cartesian theater* — referring again to Descartes (Dennett, 1991). Dennett notes that if there were a Cartesian theater, then the brain would have to perform two conversions, or *transductions*, of incoming stimuli, such as light rays or sound waves. First, it would have to convert them into patterns of neural signaling (its own internal processing language) and then it would have to convert these patterns into qualia (the medium of the show in the Cartesian theater). There would have to be a *double transduction* (Dennett, 1998).

So, the Cartesian view requires a Cartesian theater. Yet, as Dennett and others have argued, there are powerful reasons to think that there is no such thing, and that the Cartesian view is deeply flawed.

Knowing Experience

One problem for the Cartesian view is that we can have false beliefs about our own experience. For example, you might assume that your visual experience is rich and colorful right out to the edges. Yet, in fact, only the small "foveal" area in the center is like this, as you can easily prove by fixing your gaze on one point and getting someone to hold up items for identification at the edge of your visual field. You will find it hard to identify the items or say what color they are. The impression that our visual experience is uniformly rich is a sort of illusion generated by the fact that our eyes are continually darting around picking up detail here and there. We mistake the *availability* of visual information for its *presence*. Of course, this isn't an illusion in the Cartesian sense. There isn't a *second* Cartesian theater where visual experience really is uniformly rich! Rather, the illusion consists in our having false beliefs and expectations about our own experience.

This point is underlined by the phenomenon of *change blindness*. Suppose you are looking at an image on a computer monitor — say, of a city street with cars. If one of the cars were to change color suddenly, you'd expect to notice it, Yet, if the change were masked by a flicker of the screen or timed to coincide with your eye movements, then, chances are, you wouldn't. Decades of experimental research have confirmed that under such conditions viewers are often effectively blind to changes, even large ones.

Now, what would your *experience* be like in such a case? Suppose a car in the image has changed from red to green, but you haven't noticed it. The pattern of light entering your eye has changed and so has the resulting processing in the visual areas of your brain. But has the color change been reflected in the qualia show in your Cartesian theater? Suppose it has; the qualia in the relevant area are now green, not red. Then it follows that you can be wrong about your qualia, since you haven't yet noticed the color change. Suppose, then, that the change has *not* been reflected in your qualia; the qualia in that area remain red until you finally notice the change. But this puts the cart before the horse. From the Cartesian perspective, we notice colors *because of* our color qualia, not the other way round. If the qualia show is updated only *after* we have reacted to a color change, then what is the point of it?

Everyday examples reveal similar puzzles. Imagine you are looking for something in the fridge — the pickles, say. You scan the shelves but can't see the pickle jar anywhere. You turn away briefly, look back, and — lo! — there's the jar on the middle shelf. I'm sure you have had experiences like this. But now ask yourself: What was your experience of the middle shelf like *when you first looked*? What was displayed in your Cartesian theater at that location? If it was the pickle jar, then how could you have missed it? You're supposed to be directly acquainted with your qualia. But what else could have been displayed there? Some imaginary fridge item? Why would your brain have concocted such an image? A blank? That would have stood out. It's a mystery!

It seems, then, that our knowledge of our own experience isn't as secure as we think. Indeed, can you be *absolutely certain* that your qualia aren't changing all the time? Suppose that your yellow and blue qualia suddenly swap over, so that things we call "yellow" now look blue and things we call "blue" look yellow. And suppose that you also immediately *forget* how blue and yellow things used to look and assume they have always looked like this. Then you'd be unaware of the change. After all, no one else can check your qualia to confirm that they haven't changed. For all you know, this could be happening all the time!

Experience and the Brain

Another problem for the Cartesian view concerns the relation between qualia and the brain. When you have an experience, a lot of complex activity occurs in your brain, first in dedicated sensory processing regions and then across wider areas. But this brain activity is very different from the qualia you experience. You aren't directly acquainted with your own brain activity, and a neuroscientist studying your brain wouldn't be able to detect your qualia. So, it seems, qualia must be highly unusual things, which are private and distinct from the physical reality of the brain. Many contemporary philosophers accept this. Few go so far as to claim that we have nonphysical souls (as Descartes did), but they argue that consciousness is a private, nonphysical *aspect* of our brains,

which cannot be described in third-person scientific terms. (This view is known as *property dualism*, in contrast to the stronger *substance dualism* advocated by Descartes.)

But how does the brain support this nonphysical aspect? Under what conditions does sensory information acquire a nonphysical quality and enter consciousness? When and how is the second transduction achieved? This is the famous "hard problem" of consciousness. An immense amount of philosophical work has been done on it, but no consensus has emerged, and in recent years some philosophers have argued that the only solution is to suppose that all matter is conscious (a view called *panpsychism*).

A related problem concerns what *effects* this nonphysical aspect has. Once something gets into consciousness, what does it *do* there? How does the display of a yellow quale in one's Cartesian theater lead to appropriate reactions, such as saying that one sees something yellow? Dennett calls this "the Hard Question" (Dennett, 1991, p. 255). Who or what observes the display and reacts to it? If nothing does, then what's the point of it? It would be like leaving a TV playing in an empty room. If qualia are to have any effects on the brain and the behavior it controls, then they will have to be transduced *back* into patterns of neural signaling. But how is this third transduction achieved? How does the physical brain detect nonphysical qualia? If neuroscientists can't build a qualia detector, then how could evolution have done so? And how could a physical detector perform well enough to provide certain knowledge of one's qualia? It would be prone to error, just as our other sensory systems are.

Besides, if qualia have to be transduced back into neural signaling, then what's the point of them? If we wanted to record a video signal, we wouldn't play it on a monitor and film the screen; we'd plug the recorder directly into the signal source. Why wouldn't the brain do the same, linking the pre-qualia and post-qualia neurons directly, rather than creating and monitoring a qualia show? As far as effects go, everything would be the same. In short, why would we need a qualia show? The world is there for us to perceive, and we have senses to do the job. We don't need to create an internal replica of it and perceive *that*.

Consciousness as Neural Fame

These problems don't *prove* that the Cartesian view is wrong, but they do suggest that it would be worth considering an alternative. Perhaps qualia are a sort of illusion, created by the way our brains model their own activity — a view that has become known as *illusionism*. This suggestion prompts an immediate objection, however. How could qualia, of all things, be an illusion? They are the way things seem to us, and illusions are *composed* of them. If qualia themselves were illusory, then the illusion would itself require qualia — qualia of qualia. Illusionism is obviously self-defeating!

This is a natural objection, but it misses the point. Illusionists are not just suggesting that qualia are illusions; they are suggesting that we rethink the nature of

consciousness, and thus of illusions themselves. So, to understand how they respond to the objection, we need to see how that rethinking might go.

If consciousness isn't a qualia show, then what is it? Here's the general idea. For sensory information to be conscious is not for it to acquire a special *quality* but for it to have a special *effect* on us. There are many ways of developing this idea, but I'll sketch the version I prefer, which follows Dennett in likening consciousness to *fame in the brain* (Dennett, 2005).

The brain is continually probing its environment, asking what's happening, what's going to happen, and what it should be doing about it. It filters and processes the storm of incoming stimuli to extract relevant information, and it responds by making complex changes to itself and other bodily systems. Some stimuli have little effect and are soon dismissed, but others are targeted by attention mechanisms and produce waves of reaction across multiple brain systems, including those for belief, desire, emotion, memory, reasoning, decision making, and speech, as well as those that regulate bodily processes, such as heartbeat, breathing, body temperature, and digestion. (Such bodily changes are themselves monitored by internal sensory systems, producing further waves of neural changes.) Information about the stimulus has, as it were, become *famous* in the neural world, and when this happens we say that we are *aware* of the stimulus. We believe that it is there, can report its presence, and can use information about it to guide our decisions.

Now, the idea is that neural fame is conscious experience. It is worth repeating this. Neural fame *is* consciousness; it does not *produce*, or *give rise to*, consciousness, as if consciousness were something extra. The experience is simply the sum of all those internal reactions. The experience of seeing yellow, for example, is the complex of neural, psychological, and physiological reactions that occur when stimuli from yellow things achieve neural fame.

This view explains why objects produce different experiences under different observing conditions. Under unusual lighting, a blue dress might trigger the reaction pattern typically produced by yellow things. Similarly, an object may trigger different reaction patterns in different people and creatures. These differences can be large. Colorblind people have markedly different reactions to certain colors than the rest of the population do. But there are countless more subtle differences between people's reactions, reflecting their individual histories of interaction with the world. If bananas make you violently ill, this may subtly influence your reaction to yellow things in general.

The view also explains what happens when we undergo illusions and hallucinations. In such cases, the reaction pattern typical of a certain feature is triggered by something other than the feature itself. Take afterimages. If you stare at a patch of deep purple blue for a minute and then look at a white wall, the wall will initially look yellow. You will experience a negative afterimage. Your retina had adapted to purple blue light,

effectively boosting its sensitivity to the opposite color (yellow), and thus triggering the same reaction pattern a yellow surface would have done. You will think that the surface is yellow, and you will have the same yellow-related associations, emotional reactions, and so on. But what if you know that it is just an afterimage? Will you still think that the surface is yellow? In a sense, yes. The reactions I'm talking about aren't under conscious control. Your sensory systems trigger them rapidly and automatically. So, when you have a yellow afterimage, your visual system is, as it were, screaming "Yellow there!" and it will continue to do so, even if at a more reflective level you don't trust it. (By "you" I don't mean a nonphysical self; I am assuming that we are physical organisms with various levels of self-control.) You can express this by saying that the surface *seems yellow*. What you are reporting, however, is not an immediately known yellow quale, but a pattern of automatic psychological reactions to the external surface. A similar story can be told about other experiences, including bodily sensations.

The Authority of Experience Explained

But why do we treat experience reports as authoritative? Well, suppose that we have a sort of internal sense ("introspection") that enables us to recognize the reaction patterns we are undergoing. I'm not suggesting that we are directly acquainted with them, as Cartesians claim we are with qualia, just that our brains have mechanisms for monitoring their own global activity patterns, allowing us to differentiate them and classify them (for example, identifying the yellow reaction pattern as more similar to the orange one than to the blue one). As a result, we don't experience things as inert features of the world around us, but as things that are *affecting* us in some way — as attractive, repellent, scary, sinister, and so on. This is very useful for social creatures like us. For it enables us to share information about the reactions things produce in us, and so to learn from each other's experience. We don't all have to try a new fruit to know that it's bitter; someone who has tried it can tell the rest.

This explains why we treat experience reports as authoritative. They are reports of how things affect us, and no one is better placed to describe that than us. We aren't infallible, of course. We have already seen that we can make mistakes about our own experience, and it's possible that our neural self-monitoring mechanisms could malfunction. But even so, there's a way in which experience reports can't be wrong. Remember that one of the central components of the yellow reaction pattern is a disposition to think and report that the thing we're looking at is yellow, or at least *seems* yellow. Thus, in sincerely declaring that something seems yellow, we aren't just *reporting* that we are undergoing the yellow reaction pattern; we are *exhibiting* that pattern. Compare a fire alarm that, when triggered, plays a recording of someone shouting, "The fire alarm is going off!" The recording both reports that the alarm is going off and *constitutes* its going

off. The report can't be wrong because the report *is* the alarm's going off. On the view we're considering, experience reports have a similar status.

Where Are Experiences?

But where in all this are the experienced *qualities*? You might be having all the internal reactions associated with seeing a yellow object, but what about the color itself — the rich yellowy yellowness you're experiencing? The object itself isn't yellow in that way; it just has physical properties that make it reflect certain wavelengths of light. But if the yellowness is not out there in the world and not in here in your mind, then where is it?

Well, it's in neither place and both. Neither, because it isn't a simple property either of objects or of us. Both, because it is a property of objects *identified by* its effect on us. For an object to be yellow is (roughly) for it to reflect or emit light in such a way as to produce the yellow reaction pattern in most humans under typical lighting conditions. Various different physical properties might all count as yellow in this sense, provided they all have the same effect, just as many different chemical compounds all count as poisons.

Why then do we have the sense that yellow is a simple quality directly known to us? Well, because we are *also* aware of the reaction pattern yellow things produce in us, and this seems to us simple. I noted that it is useful for us to be aware of our own reaction patterns. But we don't need to know all their complex detail. We just need a sense of their overall shape, so that we can recognize them, compare them, and talk about them. So, we should expect our neural self-monitoring systems to model reactivity in a schematic way, which leaves out all the detail. We should expect to have the sense that things make a distinctive sensory impact on us, which we can't put into words, but which is immediately present and very powerful — as if it's an inner quality directly presented to a private self. And that, of course, is just what we do have. When we attend to our experience, we are in fact attending to how we are reacting internally to whatever it is we are perceiving, but because of the way our brains model this reactivity, it seems to us that we are confronted with simple, nonphysical qualities. It's a sort of illusion.

Consciousness Is Not What It Seems

Now we can answer the objection to illusionism raised earlier. An illusion isn't a qualia show; it's a pattern of reactions. Hence the illusion of *qualia* doesn't involve further qualia; it involves a further pattern of reactions. To be under the illusion of being directly acquainted with qualia is to undergo the psychological reactions direct acquaintance with qualia would produce in you, including firmly believing that you are directly acquainted with qualia. Even if there were a qualia show, it would have to have some effects; otherwise, there would be no point in it. So just imagine all those effects occurring

without an actual qualia show. Bingo! You are imagining the illusionist view of consciousness.

Your consciousness isn't a private, nonphysical world, which you know immediately and infallibly. It is a complex physical response occurring within you, which you mistake for a private, nonphysical world. It is real but not what it seems.

Suggested Reading

Daniel Dennett's work is essential reading for anyone interested in the illusionist view of consciousness. His 1988 paper "Quining Qualia" introduces a mind-bending series of thought experiments designed to undermine belief in qualia. *Consciousness Explained* (1991) mounts a sustained attack on the Cartesian theater and proposes an alternative "multiple drafts" theory. *Sweet Dreams* (2005) refines this approach, replacing the multiple drafts model with the fame-in-the brain one. The book also includes an important discussion of change blindness. (If you want to experience change blindness for yourself, there are demonstrations on J. Kevin O'Regan's website at http://nivea.psycho.univ-paris5.fr/#CBdemos.) More recent discussion of illusionism can be found in my 2017 collection *Illusionism as a Theory of Consciousness*, which contains essays by friends and foes of the view. For more on the idea that consciousness is an expressive reaction to stimulation, see the work of the neuropsychologist Nicholas Humphrey, in particular his *Seeing Red* (2006) and *Soul Dust: The Magic of Consciousness* (2011). Finally, if you are interested in Cartesian approaches to consciousness, try David Chalmers's *The Conscious Mind* (1996), which argues for property dualism, or Philip Goff's *Galileo's Error* (2019), which advocates panpsychism. For Descartes's views on foundational knowledge and the nature of mind, see his *Meditations on First Philosophy*, first published in 1641.

References

Chalmers, D. J. (1996) *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Dennett, D. C. (1988) "Quining Qualia" in A. J. Marcel and E. Bisiach (eds.) *Consciousness in Modern Science*. Oxford: Oxford University Press, pp. 42–77.

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

Dennett, D. C. (1998). "The Myth of Double Transduction" in S. R. Hameroff, A. W. Kaszniak, and A. C. Scott (Eds.), *Toward a Science of Consciousness II: The Second Tucson Discussions and Debates*. Cambridge, MA: MIT Press, pp. 97–107.

Dennett, D. C. (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. Cambridge, MA: MIT Press.

Descartes, R. (1641/1984) "Meditations on First Philosophy" in J. Cottingham, R. Stoothoff, and D. Murdoch (eds.) *The Philosophical Writings of Descartes*. Cambridge: Cambridge University Press, pp. 17–62.

Frankish, K. (ed.) (2017) *Illusionism as a Theory of Consciousness*. Exeter, United Kingdom: Imprint Academic.

Goff, P. (2019) *Galileo's Error: Foundations for a New Science of Consciousness*. New York: Pantheon Books.

Humphrey, N. (2006) *Seeing Red: A Study in Consciousness*. Cambridge, MA: Harvard University Press.

Humphrey, N. (2011) *Soul Dust: The Magic of Consciousness*. Princeton, NJ: Princeton University Press.

O'Regan, J. K. (no date) Change Blindness Demonstrations. Available at: http://nivea.psycho.univ-paris5.fr/#CBdemos (Accessed: 12 January 2025).