Philosophical Papers
(2021): 1–28

This Quintessence of Dust - *Consciousness Explained*, at Thirty

Jared Warren

Abstract: Daniel Dennett's *Consciousness Explained* is probably the most widely read book about consciousness ever written by a philosopher. Despite this, the book has had a surprisingly small influence on how most philosophers of mind view consciousness. This might be because many philosophers badly misunderstand the book. They claim it does not even attempt to explain consciousness, but instead denies its very existence. Outside of philosophy the book has had more influence, but is saddled by the same misunderstanding. Now, 30 years after publication, *Consciousness Explained* deserves reconsideration from anyone interested in consciousness. Here I make a case for this. To start, I will clear up the central misunderstanding of the book. With that done, I will explain and update Dennett's tantalizing approach to consciousness and the mind. The result brings us very, very close to explaining consciousness. Or so I will argue.

1. Dennett the Denier?

From its earliest reviews, *Consciousness Explained* (*CE*) was accused of denying the existence of consciousness. Many influential philosophers of mind made this charge in print. Here is John Searle:

> The peculiarity of Daniel Dennett's book can now be stated: he denies the existence of the data. He thinks there are no such things as … the feeling of pain. He thinks there are no such things as qualia, subjective experiences, first-person phenomena, or any of the rest of it.[1]

And Ned Block:

> In some ways, this is an extraordinary book, though *Consciousness Ignored* would be a more descriptive title.[2]

And Colin McGinn:

1 Searle 1997, page 99.
2 Block 1993, page 181.

ISSN 0556-8641 print/ISSN 1996-8523 online

DOI: 10.1080/05568641.2021.1905541
http://www.tandfonline.com

> … I don't see how [Dennett's theory] is supposed to be a theory of consciousness at all … [Dennett's theory] isn't a theory *of* consciousness … it is a disinclination to acknowledge consciousness.[3]

David Chalmers had the grace to admit that the charge is partisan:

> Dennett … spends much of his book outlining a detailed cognitive model, which he puts forward as an explanation of consciousness. … nothing in the model provides an explanation of phenomenal consciousness (although Dennett would put things differently).[4]

I could continue with similar quotes from similar sources, but you get the idea. If anything about *CE* is 'common knowledge' in philosophical circles, it is this. Nonetheless, it simply is not true.

To give discredit where it is due, Dennett himself is partially to blame for this misunderstanding. Throughout *CE*, he compares consciousness to an *illusion*, sometimes calling it a 'user illusion'. This is an old computer science term for the illusory perspective given by human–computer interfaces. The computer science-inspired 'user illusion' aspect of Dennett's theory of consciousness and the self will be discussed further below.[5] For the moment, let us focus on the base claim that consciousness itself is illusory. This is the most *prima facie* puzzling aspect of Dennett's view. Calling consciousness an 'illusion' suggests that while there *seems* to be conscious experience, in reality, there is not. Whatever else they are, illusions are cases where something seems to be the case but, in reality, is not. Yet *seeming* is itself a conscious experience. So if there *seems* to be conscious experience, to anyone or anything, then *there are conscious experiences*. This very seeming is itself one of them, so consciousness is not an illusion, QED.

Several commentators have gleefully seized on this to charge Dennett with blatant incoherence.[6] But the charge is unfair. While even the greatest

3 McGinn 1995, page 245.
4 Chalmers 1996, page 30.
5 The 'user illusion' approach to consciousness is the titular idea explored in Nørretranders 1998.
6 See Strawson 1992 and McGinn 1995, for early philosophical examples, and Harris 2014 for a popular one.

philosophers occasionally fall into inconsistency, it is difficult to imagine Dennett getting caught in so simple a trap. A minimal amount of interpretive charity requires us to assume, at least provisionally, that he has a way of avoiding this apparent inconsistency.[7]

What is it, then? Thomas Nagel recently suggested a possible answer:

> This brings us to the question of consciousness, on which Dennett holds a distinctive and openly paradoxical position. … Dennett holds that consciousness is … a particularly salient and convincing user illusion … You may well ask yourself how consciousness can be an illusion, since every illusion is itself a conscious experience … The way Dennett avoids this contradiction takes us to the heart of his position … If I understand him, this requires us to interpret ourselves behavioristically: when it seems to me that I have a subjective conscious experience, that experience is just a belief, manifested in what I am inclined to say.[8]

According to Nagel, Dennett does not think we have experiences. And when he says that we 'seem' to have experiences, or that consciousness is an 'illusion', he means only that we have certain beliefs, manifested in our experience-appropriate verbal dispositions. This offers a coherent, non-perceptual reading of Dennett's 'illusion' claims.

More carefully, for subject *S* and experience *E*, Nagel interprets Dennett as saying something like:

> There is an illusion that *S* is having experience *E* at time *t* if and only if *S* has *E*-appropriate verbal dispositions at time *t*.

Here 'E-appropriate' is a placeholder, but the basic idea is clear enough. If someone is supposedly having the experience of seeing a red square, then a disposition to say 'I am seeing a red square' or 'there is a red square in front of me' or the like, is E-appropriate. Nagel's reading is heavily influenced by Dennett's focus on verbal reports about experiences, his 'heterophemenology' (*CE*, Chapter 4). Some of Dennett's later discussions also seem to support Nagel's reading.[9] But while I understand what led Nagel to this

7 After writing this, I discovered that Dennett himself makes a similar point near the start of Dennett 2015.

8 Nagel 2017, pages 2–3.

reading, it is not the best interpretation of Dennett's position. Dennett's focus on verbal reports is *methodological*. Verbal reports are part of the data to be explained, not themselves part of an *analysis* of experience either real or illusory.

There is a more natural interpretation that avoids the contradiction. To introduce it, recall Dennett's handling of the free will problem. In his books on free will, Dennett argues for compatibilism.[10] He distinguishes different things we might mean when we call an action 'free'. Some are implausibly demanding, borderline magical—agent causation, completely uninfluenced wills, and the like. We do not have free will in anything like *that* sense, but that was not worth wanting anyway. We do have free will in a non-magical, compatibilist sense.

Dennett's approach to 'consciousness' directly mirrors this. He denies that we have conscious experiences in an implausibly demanding, borderline magical sense, but accepts that we have them in an everyday sense. When Dennett uses philosophical buzzwords for experience, like 'qualia', he balks and denies their existence (*CE*, Chapter 12). This is because 'qualia' is rich with problematic connotations concerning determinacy, indubitability, and irreducibility. As most qualia fans admit, thinking of experience in this way leads to epiphenomenalism and other dangers (*CE*, Chapters 11 and 12). So unlike Searle above, Dennett does not use 'qualia' and 'subjective experiences' as synonyms. He *almost* says exactly this several times in *CE*. For instance, after denying the possibility of zombies, using 'consciousness' in the normal way, he says:

> There is another way to address the possibility of zombies, and in some regards I think it is more satisfying. Are zombies possible? They're not just possible, they're actual. We're all zombies. [It would be an act of desperate intellectual dishonesty to quote this assertion out of context!] Nobody is conscious—not in the systematically mysterious way that supports such doctrines as epiphenomalism! I can't prove that no such sort of consciousness exists. I also can't prove

9 See especially Dennett's replies in part II of Huebner 2018.
10 Dennett 1984, 2003.

> that gremlins don't exist. The best I can do is show that there is no respectable motivation for believing in it.[11]

This distinguishes the standard notion of consciousness from the philosophers' notion. Even supposed zombies have the former, even we do not have the latter.

This ambiguity is the central cause of confusion about Dennett's basic position. Dennett does not deny the existence of consciousness, in a *lightweight* sense. He does deny it, in a *heavyweight* sense. That he does not express this with maximum clarity is probably what led some reviewers to complain about *CE*'s elusiveness and obscurity.[12] The unclarity exists because Dennett accepts 'conscious experience' in one sense, but not in another. Just as he does for 'free will'. In response, hardliners claim that if you do not accept free will in the demanding sense, then you do not accept free will in any sense worthy of the name. Dennett rejects this move in the free will debate, and he rejects the analogous move in the consciousness debate. This is his way out of the apparent contradiction.

Let us express this more precisely. The basic point can be put in a couple of different ways. One allows that we all have the same concept of *conscious experience*, but have different beliefs about the features of experience. One feature is clarity. Our visual experiences seem to involve rich, high-resolution imagery over our entire visual field. We might express this as:

> It seems that our visual experiences are perfectly clear.

Dennett accepts this, but also accepts that:

> Our visual experiences are not perfectly clear.

The earlier seeming was illusory. But there is nothing self-defeating about this; both points together can be expressed as:

> Our visual experiences initially seem to be perfectly clear, but they are not.

---

11 *CE*, page 406.

12 See Strawson 1992 and McGinn 1995.

This is consistent and coherent, since not all seemings are allowed to stand. In *CE* (mainly Chapters 3 and 11), Dennett points out that despite how it initially seems, our clarity of vision does not extend across the entire visual field. You might even call the seeming to the contrary an 'illusion'.

If this is all Dennett meant, why did he not say it directly? He sometimes did, but he was also tempted by another way of talking. This second way uses two different conceptions of experience. One is the lightweight, Dennett-friendly sense; the other is the heavyweight, Searle–Block–McGinn–Chalmers–Nagel sense. To express this unambiguously, we need two different families of terms. I will use '*D*-experience' and the like for the normal, Dennett-friendly meaning, and '*P*-experience' for the special, philosophers' meaning. We can then express Dennett's view as:

> It *D*-seems that we have *P*-experiences.

But:

> We do not have *P*-experiences.

We can gloss them together as:

> Our *D*-experiences suggest that we have *P*-experiences, but we do not.

Again, there is no contradiction once we disambiguate.

I think this is the right interpretation of Dennett's apparent denials of consciousness. The clearest statement I have found from Dennett is in an interview published in 2006:

> People don't like me saying that they're not conscious of as much as they thought they are, and what they are conscious of doesn't have the features that they say it has. Their reaction to this is 'Oh Dan's just denying the existence of consciousness'. No, I'm not. I'm just saying that it's not what they thought it was.[13]

13 Quoted in Blackmore 2006, page 85.

One would hope this would be the end of the matter, and that people would stop calling Dennett a 'consciousness denier'. But while hope springs eternal, recent history makes me pessimistic. Recently, Dennett's 'illusion' description has been taken up and elaborated by others.[14] I have no gripe with the content of this, provided it is cashed out in one of the two ways just explained, but the 'illusion' terminology is apt to cause misunderstandings and should probably be avoided. As evidence for this, I point to *30 years* of *CE* being misunderstood.

Dennett's view is not directly self-contradictory, but it might have other problems. To assess it, we need to move beyond slogans and into the details.

2. Multiple Drafts and Fame in the Brain

In *CE*, Dennett calls his theory of consciousness the *multiple drafts theory*. This metaphor never really caught on, and he later abandoned it in favor of another—*fame in the brain*, or *cerebral celebrity*. This change of name did not amount to a change of doctrine, but it did involve a slight change of emphasis.

Dennett introduces his positive theory by contrasting it with *Cartesian materialism*. This is any materialist theory that appeals, explicitly or implicitly, to a *Cartesian theater*—a *place* in the brain where items enter into consciousness at *precise times*, where they are then and there available to an *inner observer*. Dennett presents a lot of this metaphorically. Although he never characterizes Cartesian materialism precisely, he generally assumes that it involves a special place, a special time, and an inner observer.[15]

Many critics have argued that Dennett was tilting at windmills. These critics claim that Cartesian materialism is not and has never been widely endorsed. I do not think this is right, but by presenting his central target metaphorically Dennett practically asked to be misunderstood. I think the theoretical point of Dennett's attack on Cartesian materialism is twofold. First, it allows him to argue for a certain indeterminacy claim.

14 See many of the essays in Frankish 2017, for example.
15 Schneider 2007 offers a similar reading.

Second, it allows him to introduce a way of conceptualizing the brain's consciousness-relevant architecture without a Cartesian theater. Let us take each of these points in turn, starting with indeterminacy.

Suppose you have a misleading visual experience of seeing a glasses-free woman on a street corner *as* a different, glasses-adorned woman you actually met at a party last week. Dennett distinguishes two accounts of what went wrong:

ORWELLIAN: You had an initial, brief moment of experiencing the woman glasses-free, but your memory of the glasses-adorned woman contaminated it almost immediately, and the earlier experience left no trace in memory. (Named for the rewriting of historical records in Orwell's *Nineteen Eighty-Four*.)

STALINESQUE: You never experienced the woman as glasses-free, instead your memory of the glasses-adorned woman contaminated information about the glasses-free woman on its way to being experienced. (Named for Joseph Stalin's phony show-trials.)

Dennett claims (*CE*, Chapter 5) that when the time scale is small enough, *there is no factual difference between these two accounts*:

> … the distinction between pre-experiential revisions that *change that which was experienced* and post-experiential revisions that have the effect of *misreporting or misrecording what was experienced* is indeterminate in the limit.[16]

Rejecting the Cartesian theater forces this upon us—spatial smearing in the brain leads to temporal smearing, so that at small enough timescales, the facts give out. In the other direction, direct arguments for temporal smearing undermine the very idea of a 'finish line', which is required for the Cartesian theater.

16 *CE*, page 247.

Dennett's most direct argument for temporal indeterminacy in *CE* admittedly veers toward verificationism. This has led to a standard reply: Not so fast. We should be open to indirect ways that evidence can bear on a question.[17] Sometimes indirect evidence can even be conclusive. This is a central lesson of twentieth-century philosophy of science. Even if verificationism is true, empirical science is so complex that we should be hesitant to declare, in advance, that a putative scientific question has no determinate answer. When we offer arguments for indeterminacy, we should do so cautiously.

This reply has *some* merit against *some* of Dennett's arguments on this point. But elsewhere, Dennett has given more detailed arguments that do not assume a strong form of verificationism.[18] So a rejection of verification does not undercut the truly interesting philosophical point being made. Rhetorically, Dennett's close focus on this particular example (*CE*, Chapter 6) might have been strategically unwise. The main point of interest is *the very idea* that there could be indeterminacy about whether our experience over time period t was E_1 as opposed to E_2, where E_1 and E_2 are incompatible. Fans of heavyweight conceptions of consciousness will sometimes admit the possibility of indeterminacy, if pressed, but you rarely get the sense that they take it seriously. By alerting us to this possibility and making it vivid, Dennett performed an important service.

His discussion shows that almost any broadly *functional* account of consciousness must take indeterminacy seriously. There might be no functional difference at all, in the consciousness-relevant sense, between having potential experience E_1 as opposed to potential experience E_2. In such cases, while it might be determinate that *some* experience is being had, there is no fact about whether it is *this one* as opposed to *that one*. You might find this strange. It gets stranger still. After admitting this possibility, we soon

17 Block 1993 replies in something like this manner.

18 In most detail in Dennett and Kinsbourne 1992. And more recently Dennett (2015) has connected his arguments on these points with currently popular predictive models of neurocomputation, like the one advocated in Clark 2013.

realize that we should also admit deeper types of potential indeterminacy—including indeterminacy about whether the global mental state of a creature at time t is conscious or not. And this leads to the possibility of creatures whose only consciousness-candidate states reside in these borderlands. For such creatures, it is indeterminate whether they are conscious at all.

Dennett's focus on the Orwellian/Stalinesque distinction sometimes obscures this, but in *CE* he also endorses the possibility of these more radical types of indeterminacy. Most philosophers of mind find this ridiculous. Several have even argued against indeterminacy, at length.[19] But their arguments start from premises about consciousness that Dennett (rightly) rejects. Dennett's indeterminacy claims, more than anything else in *CE*, help to wake us from our dogmatic slumbers. Merely entertaining them as serious possibilities starts us on the path to thinking about consciousness in a more serious, less magical fashion.

This brings us to the second philosophical point behind his attack on the Cartesian theater—the alternative, indeterminacy-friendly architecture. The human brain is a biological neural network. It works in a massively parallel and distributed fashion. This is not to deny localization for particular functions. So far this is common ground, close to established scientific fact. Almost all cognitive scientists agree that, in some way, a brain generates a mind via the operation of nested, interacting computational subsystems. Terms like 'demons', 'agents', and 'modules' are used for these subsystems, although sometimes the specific terms import additional, more tendentious, assumptions (hardwired, single-function, informationally isolated, and so on). I will call these subsystems 'demons' without making any additional assumptions about them.

The problem of getting from this basic functional-computational picture to anything like a *conscious* mind is what motivates Cartesian materialism. Jerry Fodor provides a characteristically vivid statement of this:

19 See Antony 2006 and 2008, Simon 2017, and O'Rourke n.d.

> Eventually the mind has to integrate the results of all those modular computations, and I don't see how there could be a module for doing *that*. If, in short, there is a community of computers living in my head, there had also better be somebody who is in charge; and, by God, it had better be me.[20]

Dennett agrees that there is no special consciousness center in the brain, no special version of the neural code transduction into which constitutes consciousness. The problem is in accepting this while also retaining the naturalist assumption that the functional-computational approach in cognitive science is *metaphysically complete*. Dennett's alternative architecture is his solution to this problem.

Instead of appealing to an inner observer, we must look to interactions *between* brain systems to understand consciousness. Dennett was not the first to have this insight, and like a number of theorists before him, he was influenced by Oliver Selfridge's early 'pandemonium' architecture.[21] This was one of the earliest non-top-down computational approaches, and perhaps the first that could learn. Roughly, a pandemonium involves computational demons nested in a quasi-hierarchical structure. Demons (or coalitions of demons) at each level compete with each other. The one that 'shouts the loudest' gets its message picked up at the next level, and so on. Pandemonium was originally applied to visual pattern matching, but in the very first public discussion of pandemonium, John McCarthy suggested using it to explain consciousness:

> I would like to speak briefly about the advantages of the pandemonium model as an actual model of conscious behavior. In observing a brain, one should make a distinction between that aspect of the behavior which is available consciously, and those behaviors, no doubt equally important, but which proceed unconsciously. If one conceives of the brain as a pandemonium—a collection of demons—perhaps what is going on within the demons can be regarded as the unconscious part of thought, and what the demons are publicly shouting for each other to hear, as the conscious part of thought.[22]

---

20 Fodor 1998, page 207.
21 Selfridge 1959.
22 McCarthy 1959, page 147.

A more detailed pandemonium model of consciousness was suggested by John Jackson in the 1980s.[23] Marvin Minsky's similar idea, that the mind consists of a 'society' of mindless computational 'agents', appeared at around the same time.[24] Dennett's approach is directly in line with these ideas.

Key to all of them is the rejection of the Cartesian theater. There is no inner observer, there is only 'shouting' from demons to other demons. A shout becomes conscious when it reaches a level of system-wide influence. What are the demons shouting? There is no standard term—'information', 'messages', 'signals', 'content', and 'representations' have all been used. I will usually use 'information' in an attempt to stay non-committal.

The 'multiple drafts' metaphor is based on the thought that when an organism takes in sensory information (for example), various demons start operating on that information, in parallel, repackaging, reworking, and adding to it in their own particular ways. No specific demon is the final 'publishing' mechanism. Dennett agrees with Fodor that it is hard to imagine a demon for doing *that*. Instead, revisions and alterations are made repeatedly, in a diachronic, ongoing, parallel fashion. This metaphor pushes us toward recognizing Orwellian/Stalinesque temporal indeterminacy, but not directly toward other types of indeterminacy. It also does not hint at what is criterial of consciousness.

Dennett later replaced it with a metaphor that did—'fame in the brain' or 'cerebral celebrity'.[25] Something enters consciousness when it gains system-wide influence. This consists in being broadcast to and received by other systems—written down in memory, made available for reasoning processes, fed into planning and behavioral systems. Which signals become famous is partly probe-dependent (which is why worrying about a hitherto unnoticed spot on your leg soon makes it feel like something is touching it).

23 Jackson 1987.
24 Minsky 1986.
25 See Dennett 2001.

As with fame, there is no precise threshold. Is Daniel Dennett famous? Maybe, but not as clearly as Tom Cruise is. The line is blurry. The distinction admits of borderline cases. So too with consciousness. Some critics have been baffled by this, perhaps because they mistake it for a much stronger claim. Here is Colin McGinn:

> What Dennett needs to say—and I am still not sure that he is willing to say it—is that among all the informational states present in the brain at a given time there is just no distinction, firm or loose, between those that reach consciousness and those that do not. This strikes me as obviously false empirically ….[26]

Indeed this is obviously false empirically, but it is also not something that Dennett needs to say. Dennett is not arguing that there is *no distinction*, firm or loose, between states that reach consciousness and those that do not. Instead he is saying there is no firm distinction, only a loose one—or better, a *fuzzy* one. This is compatible with there being clear cases on both sides of the distinction. Some shouts are clearly conscious, others clearly unconscious, while still others reside in the borderlands.

This 'multiple drafts' account is not yet a full theory of consciousness. Instead, it sets up Dennett's theory. The most important parts of the setup are those I have detailed: (*i*) consciousness can be indeterminate in various ways; and (*ii*) the consciousness-enabling architecture of the mind does not include anything like an inner observer, instead something becomes conscious when it reaches system-wide influence. What is the argumentative relationship between these two points? It depends. You can start with indeterminacy and then look for a model of consciousness that allows it. Or you can start with a model and then argue to the possibility of indeterminacy. One strategy may be appropriate in one dialectical context, the other in another, but the two components of Dennett's non-traditional approach are self-reinforcing. This is not circularity, it is internal coherence.

---

26 McGinn 1995, page 245.

3. Joycean Machines, Global Workspaces, and Broadcasting

If information becomes conscious by gaining system-wide influence, the obvious question is: *How does that happen?* A precise neurocomputational model is too much to ask for at this point, but something more should be said about the processes involved. There are also immediate challenges, since Dennett's account thus far sits ill with the existence of our familiar, subjectively experienced *stream of consciousness*.

Almost the entire time we are awake, we have a serial flow of experiences, one after another. The seriality of experience is both obvious and empirically well confirmed. Of course, this does not mean that you cannot hear a sound at the same time you see an image. You can, you do. It means that there are competing tracks from the same sensory modality, and that even different sensory modalities generally integrate into a coherent whole. The consequences of this are vivid. You can see the figure as a duck, then as a rabbit, but you cannot see it as both at once. The necker cube is oriented one way, then the other, not both at once. Even if this breaks down in the temporal limit, at a coarser grain it is difficult to deny. Given that the brain is a parallel, distributed processor with many different things happening at the same time in different places, why is experience serial?

In *CE*, Dennett recognizes the difficulties of accounting for seriality with 'multiple drafts', so he adds another layer to his theory (Chapters 7, 8, and 9). He claims that the biological, parallel computer that is the human brain, our hardware, implements a virtual, serial computer. In saying this, he draws on the computer science notion of a *virtual machine*. Your home computer was not hardwired to be a word processor, but it can 'simulate' a word processor. In the last 30 years the term has lapsed; it is now more common to talk instead about *software* or *applications*.

At one level, Dennett's move is quite natural. There is no denying that language, mathematics, diagrams, and other cultural products enrich our powers of thought. The human brain has not changed much, biologically, in the last 10,000 years, yet our cognitive powers have exploded in that time. Dennett thinks that the products of culture install a serial virtual machine—

the Joycean machine (named for James Joyce's stream-of-consciousness passages in *Ulysses*)—on our parallel wetware. This also fleshes out Dennett's 'user illusion' comparison, mentioned above. He appeals to a meme-based perspective to explain the sense in which culture 'installs' this abstract machine in our heads.[27] At the moment, what matters is not *how* the Joycean machine gets installed, but *what it can do* once it is. The Joycean machine allows for various bits of information to attain fame in the brain in a way that accounts for the serial nature of the apparent stream of consciousness, while still allowing for a healthy amount of indeterminacy.

The idea of a 'virtual machine', and Dennett's use of it, might give us pause. To say that something is 'virtual' suggests that it is not 'genuine' or 'real'. But at this point in history, if you told someone that the word processor on their laptop was not 'real', they would look at you the same way they would if you claimed to be a poached egg. So-called 'virtual' machines are *real*, they just are not *real machines*. The states of the word processor on your laptop do not correspond, in an obvious way, to the physcial states of the laptop's macroscopic components. By this, I am obviously *not* denying that the physical workings of the laptop implement the word processor—I am not a word processor *dualist*! I am simply saying that the word processor exists at the level of software rather than hardware.

So Dennett is not denying that the Joycean machine *exists* by calling it 'virtual'. Yet even with this clarification made, questions remain. With designed computing devices like desktop computers, laptops, and smartphones, the distinction between hardware and software is clear. Yet it is not nearly as clear when it comes to *brains*, at least in detail. In broad outline, though, things are uncontroversial. The 'hardware' of a brain consists mainly of a densely interconnected web of neurons in the skull. And the 'software' of a brain includes the various computational demons

27 This is stressed in Dennett 2017. The approach of Heyes 2018 might also be adopted. A lot of what Dennett says about the Joycean machine does not strictly *require* a memetic theory of how the machine gets installed.

discussed above. Things become a bit hazy at this point, because some of the demons come pre-installed in animal brains. And even some that are not pre-installed are installed during the normal developmental process.

For Dennett's purposes, and ours, the haziness can be sidestepped. The Joycean machine is a software description of the brain *as a whole*, corresponding to the conscious mind's stream of consciousness. It is serial, transitioning from state to state. And its states, though implemented in the operations of the physical brain, do not correspond to obvious, large-scale physical operations of the brain. Of course, philosophical challenges remain for the idea that software or programs are implemented by physical systems, like brains, but these challenges are not specifically targeted at Dennett's approach, so I will not pursue them here.[28] If there is a reasonable naturalistic theory of computational implementation, it should serve Dennett's needs.

The Joycean machine is not an irrelevant add-on; it is instead the heart of Dennett's theory of consciousness:

> … I hereby declare that YES, my theory is a theory of consciousness. Anyone or anything that has such a virtual machine as its control system is conscious in the fullest sense, and it is conscious *because* it has such a virtual machine.[29]

Despite this proclamation, it is not immediately clear how the virtual machine portion of Dennett's theory relates to the multiple drafts portion. He hints at two related ways of bringing them together.

The first hint comes after a discussion of the language processing that goes on in the virtual machine:

> … these activities magnify and transform the underlying hardware powers in ways that seem (from the "outside") quite magical. But still (I am sure you want to object): All this has little or nothing to do with consciousness! After all, a von Neumann machine is entirely unconscious; why should implementing it—or something like it: a Joycean machine—be any more conscious? I do have an answer: The von Neumann machine, by being wired up from the outset that

28 The challenges I have in mind take off from the appendix to Putnam 1988.
29 *CE*, page 281.

> way, with maximally efficient informational links, didn't have to become the object of its own elaborate perceptual systems. The workings of the Joycean machine, on the other hand, are just as "visible" and "audible" to it as any of the things in the external world that it is designed to perceive—for the simple reason that they have much of the same perceptual machinery focused on them.[30]

Here it sounds like Dennett wants to explain consciousness in terms of self-representation. Indeed, later in *CE* (Chapter 10) he sympathetically discusses David Rosenthal's higher-order thought approach to consciousness, but stops just short of endorsing it.[31] Still, Dennett clearly sees self-monitoring and higher-order thought as key parts of human experience, as partly constitutive of self-consciousness, and as crucial to the creation of the 'self' (*CE*, Chapter 13).[32]

The second hint ties the Joycean machine directly to consciousness, not just self-consciousness. Dennett suggests that the powers of the Joycean machine are required for *anything* to gain system-wide influence—that is, to be consciously experienced. Indeed, on page 280 of *CE*, Dennett explicitly describes the Joycean machine as implanting 'an internal global workspace in which demons broadcast messages to demons, forming coalitions and all the rest'. This ties Dennett's view directly to subsequently popular 'global workspace' approaches, except Dennett thinks that creatures can only have a global workspace if they are running a Joycean machine. Since the Joycean machine requires cultural tools—principally language—this entails that, for Dennett, *non-linguistic creatures do not have conscious experiences.*

Most of us find this implausible. Are not *some* non-linguistic animals conscious? Fortunately, there is a way to revise a Dennett-style position,

30 *CE*, pages 225–226.

31 See Rosenthal 1990.

32 Some recent theories of the self, self-consciousness, and consciousness have important aspects in common with Dennett's approach, even—in some cases—adopting the 'user illusion' metaphor. For examples, see Humphrey 2006, 2011, Metzinger 2003, 2009, Nørretranders 1998, Prinz 2012, and Seth 2021, as well as the global neuronal workspace approaches cited below.

extending and amending it, but retaining its most important insights. To start, we should allow for the possibility of brain-wide broadcasting systems *built by evolution*. This is any biological mechanism that allows shouts from demons or coalitions of demons to be broadcast system-wide, so as to achieve 'fame in the brain'. This move very closely aligns Dennett's approach with standard 'global workspace' (*GW*) theories.

The basic global workspace idea, under that name, was invented by Bernard Baars.[33] It was later developed into global neuronal workspace approaches by Stanislas Dehaene, Lionel Naccache and others.[34] *GW* theory agrees with Dennett's 'fame in the brain' idea. The role of conscious experience is to integrate, provide access to, and coordinate the functioning of various specialized demons and coalitions. Most *GW* theories see consciousness as a briefly active event in working memory that corresponds to what we subjectively experience. Like the Joycean machine, the *GW* works serially over unconscious parallel processes in the brain. Given all of this, it is easy to see why Dennett nearly identified his 'multiple drafts' approach as a version of *GW* theory in at least one later discussion.[35] But the exact connection between *GW* theories and Dennett's theory is unclear. To start, there is a serious source of *prima facie* tension between them.

Baars often used an 'inner theater' metaphor to explain global workspace theory, coming perilously close to the Cartesian theater. This has led some commentators to argue that Dennett's theory cannot possibly align with *GW* theories.[36] But I think Dennett's approach can be thought of as a global workspace theory *without the workspace*. This is cheeky. What I mean is that the workspace is not really a place, or a special system; it is instead a *functional* notion. So we can best make sense of the insights of global workspace approaches by dropping the 'theater' metaphor entirely,

33 See Baars 1988, 1997.
34 See Dehaene and Naccache 2001 for an overview and Dahaene 2014 for a comprehensive popular summary.
35 See Dennett 2001.
36 See Schneider 2007, for example.

putting less emphasis on the 'workspace' idea and focusing on the 'broadcasting' idea that *GW* theorists also appeal to.[37]

Both Dennett and *GW* theorists think that consciousness is a matter of system-wide influence. There is no central executive. When demons or coalitions of demons shout loudly enough for the entire system to hear, those shouts are conscious. The difference is that Dennett thinks only brains running culturally designed Joycean machines generate conscious experiences while *GW* theorists think that some simpler animal brains have biologically built software for facilitating system-wide broadcasting, using working memory and attentional mechanisms. Given the empirical success of *GW*/broadcasting models, the extreme version of Dennett's view should be abandoned. Some but not all animal brains have *biological* broadcasting systems. Creatures with these biological broadcasting systems have conscious experiences and something like a stream of consciousness.

Yet we should not scrap the Joycean machine. We should instead combine these ideas, and see the pre-existing biological broadcasting system as what enables easy installation of the cultural Joycean machine. I do not think this conflicts with Dennett's memetic story of the Joycean machine's installation in the brain. If anything, it better explains *how* Dennett's memetic story—or other installation mechanisms—could work. The types of behavior and planning enabled by biological *GW* broadcasting allow animals to acquire the complex abilities that constitute the Joycean machine. So, Dennett's point about powers is well made. Our cultural tools, especially language, allow us to *upgrade* the *GW* broadcasting system. Our cultural tools also allow the 'installation' of new, specialized demons, not hardwired by evolution.

This is not simply the possibility of *learning*. These tools also allow different bits of information to be shouted by brain demons, new and old, while

37 I subsequently discovered that Godfrey-Smith 2016 makes a similar suggestion; see pages 149–151. He also notes that in this context 'broadcasting' may not even be a metaphor, but instead the literal truth.

also allowing for better and more efficient internal signal broadcasting. This is the import of Dennett's oft-made points about the power of self-talk and other kinds of self-stimulation (*CE*, Chapters 7 and 8). It dovetails with and extends what *GW* theorists say about why consciousness is such a valuable thing for a creature to have. To simplify only slightly, conscious creatures, but not unconscious creatures, can engage in long-term planning, general reasoning, and non-local pattern recognition. They can engage in so-called 'system 2' thinking that is unavailable to unconscious creatures.[38] The Joycean machine not only enables richer and more powerful types of system 2 thinking, it also enables entirely new kinds of experiences.

Non-linguistic creatures can have conscious experiences. But some conscious experiences *require* a Joycean machine. There are levels of consciousness, in a sense. In relatively simple creatures, perhaps only a few deeply important signals get broadcast system-wide. Fish might get by without a stream of consciousness, with only occasional signals—*danger*!, *pain*!—attaining system-wide influence. If so, then usually there is nothing it is like to be a fish, but then occasionally there is something it is like, and it sucks. In more sophisticated creatures, there is probably something *like* a serial stream of consciousness. Many mammals probably fall into this category.

Yet we should not be cavalier. Empirical evidence suggests that complex biological broadcasting systems might result in something like the constancy and coherence of a human stream of consciousness.[39] But we should admit not only that the Joycean machine can broadcast *different* and *richer* information, but also that Joycean machines, with their capacity for self-talk, also increase the *coherence* of broadcasting. Even the most robust biological broadcasting systems fall short of the powers of the Joycean machine. This leaves most or all non-human animals with, from our perspective, limited experiences. They can feel pain, but they cannot

38 See Kahneman 2011 for the system 1/system 2 distinction.
39 For example, see the evidence reported in Chapter 7 of Dahaene 2014.

suffer.[40] They can be sad, but they cannot feel despair. The Joycean machine enables a wider and richer range of conscious experiences, to the point where we might even say that Joycean consciousness is different in character and kind from non-Joycean consciousness.

We should all admit that 'what it's like' to be a human with a Joycean machine is extremely different from 'what it's like' to be a chimpanzee, without one. Despite the biological similarities, the differences in conscious experience are vast. This idea harmonizes Dennett's 'fame in the brain' characterization with his Joycean machine theory without leading to unpalatable verdicts about which non-human creatures are conscious.

Of course, hooking up 30 old Dell computers from the 1990s and allowing them to share information probably would not result in conscious experiences of any kind. Arguably, we need roving systems that are taking in rich sensory information about the world and using said information to generate behavior. Conscious experience results when a global broadcasting mechanism operates as the 'executive control' in such systems. Recognizing this much of a connection between consciousness and behavior does not require 'behaviorism' or any other objectionable philosophical doctrine.

So, a new metaphor. A large building with rooms and hallways. Within, demons fight and cooperate and yell to each other. In some buildings, only very loud shrieks reverberate throughout the entire building, and this only happens occasionally. These correspond to things like pain and danger signals in simple creatures. In other, more complex buildings, the acoustics are improved. More and more shouts get broadcast. Some of these buildings are even set up so that there is almost always some coherent message being broadcast. These correspond to more complex creatures, with something like a global workspace or a complex biological broadcasting system. Finally, we have the most sophisticated buildings of all. These buildings might be nearly the same size and layout as others, but they are fitted with an intercom system. This allows for better message fidelity and for more

40 See Chapter 6 of Dennett 1996.

complex messages to be broadcast and to be broadcast more coherently. These buildings correspond to creatures running Joycean machines. To us.

Call this a *multi-level, global broadcasting theory of consciousness*. It is like a pluralistic, leveled global workspace theory, but where the 'workspace' is understood in purely *functional* terms. This is important. Some global neuronal workspace theorists have been tempted to precisify our concept of *consciousness* by, in effect, replacing it with a more precise *neuro-biological* notion, appealing to the neuro-biological realization of the global workspace. But this move threatens to reintroduce the Cartesian theater along with all of the old misunderstandings that accompanied it. System-wide broadcasting is a functional notion. Only by understanding it as such can we retain all of the interesting and important conclusions about indeterminacy discussed above. The functional notion of system-wide influence comes in degrees, and thus, so too does consciousness.

Different types of biologically and culturally installed broadcasting software are being run on different, biological brains. But in all systems, consciousness is explained in terms of system-wide broadcasting and receiving. In organic creatures like us, this is just fame in the brain.

4. Philosophical Q&A

Time for some quick questions and answers, for explicitness, clarity, and to see where we are in terms of explaining consciousness. The answers I give have not all been *explicitly* given by Dennett, and some appeal to my multi-level global broadcasting modification of Dennett's theory. Still, I think almost all of them find direct or indirect support in *Consciousness Explained* and Dennett's more recent writings on consciousness.[41] I present them here baldly and boldly, so get ready for some provocation.

When is something consciously experienced by a creature? When it reaches system- wide influence. Nothing else needs to happen; this is what it is for something to be conscious.

41 In addition to *CE*, see especially Dennett 1988, 2001, 2015, and Cohen and Dennett 2011.

How does anything attain this influence? By system-wide broadcasting, either using a biological broadcasting system, built by evolution (so to speak), or a culturally upgraded system, when language, diagrams, and other cultural tools have upgraded the system software, enabling non-hard-wired demons along with new broadcasting channels.

Does it account for self-consciousness? Yes, by way of the self-reflexive representational capacities of some creatures, especially those with Joycean machines. Perhaps also those with sophisticated biological broadcasting systems.

Does it account for the stream of consciousness? Yes, Joycean machines and sophisticated biological broadcasting systems operate in a constant, roughly serial fashion over many demons working in parallel. Simpler creatures may occasionally have experiences, but lack a stream of consciousness.

Does consciousness come in degrees? Yes. There is no Cartesian theater, and even the global workspace should be understood functionally. It can be indeterminate whether the experience a creature is having is E_1 or E_2. It can also be indeterminate whether something has achieved the level of influence required for consciousness. For some creatures, *all* of their consciousness candidate states may be in the borderlands. It is indeterminate whether such creatures are conscious at all.

Are there different kinds of consciousness? Yes, in a sense. Biological broadcasting systems come in various degrees of complexity. And culturally upgraded systems are more complex still. Perhaps from here to infinity. Different broadcasting systems have different powers, and not every creature's demons can shout the same messages. Human experience is richer than non-linguistic animal experience. In some contexts it might even be useful to distinguish these as two different types or kinds or (to order them) *levels* of conscious experience.

Is this a theory of 'what-it's-like-ness'? Yes. When the shouts of internal demons in a creature of some complexity reach system-wide influence, there is something it is like to be that creature, at that time. But talk of 'what-it's-like-ness' adds little over and above talk of 'consciousness' or 'experience', save for a greater risk of confusion. For instance, it pushes us to ask questions like … .

Can we know what it is like to be a bat? Bats are probably conscious, so by definition, there is probably something it is like to be a bat. Yet since a bat's broadcasting system and the signals it can broadcast are very different from our own, we might say that we do not 'know' what it's like to be a bat. But then again, *neither do bats*! Non-trivial 'knowledge' of what's it's like to be something likely requires the kind of higher-order reflection enabled by a Joycean machine.

Is this a version of functionalism? Yes, it is a version of computational functionalism. Any signal that plays the right computational-functional role, of having system-wide influence, is conscious.

Is it a version of the identity theory? No, in principle, systems with consciousness-generating architectures can be built out of anything. Although, in reality, all conscious experiences we are aware of are generated by biological brains. There may even be engineering constraints on which types of things can be used to build conscious minds. So, multiple realizability is accepted in principle, but whether it happens in practice is an empirical question.

Is it a version of behaviorism? No. The same behaviors could, in principle, be generated via a gigantic look-up table, or by way of a much, much *wider* and more *finely tuned* but less *sophisticated* neural network like OpenAI's GPT-3 (98 layers and 175 billion parameters, with a training data set comprising the majority of the internet). Creatures with control systems like these will not be conscious, even if they are behaviorally identical to other, conscious creatures.

What, then, is the metaphysical nature of consciousness? In philosopher-speak, it is a version of analytic computational functionalism. There is no gap between something attaining fame in the brain and it being conscious. To think there is a gap between these is a mistake, akin to asking, 'That is a three-sided polygon, but is it a triangle?'—to seriously ask about a gap here amounts to misunderstanding the theory.

What about qualia, P-consciousness, intrinsic subjective character, and the like? Find the nearest wastebasket.

Why is anything conscious? Consciousness performs an important role. It allows demons and coalitions of demons to coordinate and integrate

information without any homunculus. Conscious creatures can do things that most unconscious creatures cannot do.

But could evolution not have done all of this without creating consciousness? Not really. The behavioral isomorphs mentioned above are not anything that evolution could build—except indirectly, by first building minds like ours. There is not world enough, nor time, to *directly* build and train a GPT-3, or a gigantic look-up table. Building a conscious creature is the easiest way to get the kind of online control that helps complex, roving organisms get by and thrive in the world. And, once again, to think that the same computational architecture and the same processing could take place without any lights being on is simply to reject the theory, not to offer an independent objection.

So, has consciousness been explained? Not completely. We are still building and refining more and more detailed and accurate neurocomputational models. Eventually, though perhaps not for many years, these models will be complete and exhaustive. When they are, the *exact* computational mechanisms of consciousness will be known and understood.

Does any philosophical work remain? As the scientific understanding of consciousness advances, many small philosophical puzzles will pop up, so there will be many things for philosophers of neuroscience and cognitive science to busy themselves with. But the important big-picture developments will be further along on the trail blazed by Dennett. Some theoretical work in this direction remains to be done, but the *shape* of the overall philosophical theory of consciousness is coming into view. At this point, the preparatory work is all but finished. With hindsight and some added refinements, Dennett's *Consciousness Explained* can be seen as a crucial part of that exciting endeavor. So, in an important sense, consciousness has been explained, philosophically. Puzzles still abound, but remaining mysteries are few.[42]

42 Thanks to Rosa Cao, Daniel Dennett, Philippe Lusson, and Douglas Stalker.

References

Antony, Michael V. (2006). 'Vagueness and the Metaphysics of Consciousness'. *Philosophical Studies* 128: 515–538.

———. (2008). 'Are Our Concepts *Conscious State* and *Conscious Creature* Vague?' *Erkenntnis* 68(2): 239–263.

Baars, Bernard. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press: Cambridge.

———. (1997). *In the Theater of Consciousness: The Workspace of the Mind*. Oxford University Press: Oxford.

Blackmore, Susan. (2006). *Conversations* on *Consciousness*. Oxford University Press: Oxford.

Block, Ned. (1993). 'Review of Daniel Dennett, *Consciousness Explained*'. *Journal of Philosophy* 90(4): 181–193.

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

Clark, Andy. (2013). 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science'. *Behavioral and Brain Sciences* 36(3): 181–253.

Cohen, Michael A. and Daniel C. Dennett. (2011). 'Consciousness Cannot Be Separated from Function'. *Trends in Cognitive Science* 15(8): 358–364.

Dehaene, Stanislas. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Penguin: New York.

Dehaene, Stanislas and Lionel Naccache. (2001). 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework'. *Cognition* 79: 1–37.

Dennett, Daniel C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. The MIT Press: Cambridge.

———. (1988). 'Quining Qualia'. In Marcel and Bisiach (eds), *Consciousness in Modern Science*. Oxford University Press: Oxford: 42–77.

———. (1991). *Consciousness Explained*. Back Bay Books: New York.

———. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. Basic Books: New York.

———. (2001). 'Are We Explaining Consciousness Yet?' *Cognition* 79: 221–237.

———. (2003). *Freedom Evolves*. Penguin Books: London.

———. (2015). 'Why and How Does Consciousness Seem the Way It Seems?' In Metzinger T. and J.M. Windt (eds), *Open MIND 10(T)*. Mind Group: Frankfurt am Mein.

———. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. Back Bay Books: New York: 1–11.

Dennett, Daniel C. and Marcel Kinsbourne. (1992). 'Time and the Observer: The Where and When of Consciousness in the Brain'. *Behavioral and Brain Sciences* 15(2): 183–201.

Fodor, Jerry. (1998). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. The MIT Press: Cambridge.

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

Godfrey-Smith, Peter. (2016). *Other Minds: The Octopus, The Sea, and The Deep Origins of Consciousness*. Farrar, Straus and Giroux: New York.

Harris, Sam. (2014). *Waking Up: A Guide to Spirituality Without Religion*. Simon & Schuster: New York.

Heyes, Cecilia. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University Press: Cambridge, MA.

Huebner, Bryce (ed.). (2018). *The Philosophy of Daniel Dennett*. Oxford University Press: Oxford.

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

———. (2011). *Soul Dust: The Magic of Consciousness*. Princeton University Press: Princeton.

Jackson, John V. (1987). 'Idea for a Mind'. *SIGART Bulletin* 101: 23–26.

Kahneman, Daniel. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux: New York.

McCarthy, John. (1959). 'Discussion of Oliver Selfridge, "Pandemonium: A Paradigm for Learning"'. In *Symposium on the Mechanization of Thought Processes*. H.M. Stationery Office: London: 527–531.

McGinn, Colin. (1995). 'Consciousness Evaded: Comments on Dennett'. *Philosophical Perspectives 9: AI, Connectionism and Philosophical Psychology*: 241–249.

Metzinger, Thomas. (2003). *Being No One: The Self-Model Theory of Subjectivity*. The MIT Press: Cambridge.

———. (2009). *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books: New York.

Minsky, Marvin. (1986). *The Society of Mind*. Simon & Schuster: New York.

Nagel, Thomas. (2017). 'Is Consciousness an Illusion?' *The New York Review* March 9: 32–34.

Nørretranders, Tor. (1998). *The User Illusion: Cutting Consciousness Down to Size*. Penguin Books: New York.

O'Rourke, Joshua. (n.d.). 'Phenomenal Consciousness Must Be Sharp'. Unpublished Manuscript.

Putnam, Hilary. (1988). *Representation and Reality*. The MIT Press: Cambridge, MA.

Prinz, Jessie. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press: Oxford.

Rosenthal, David. (1990). 'A Theory of Consciousness'. In *ZIF Report 40*. Bielefield University: Germany.

Schneider, Susan. (2007). 'Daniel Dennett on the Nature of Consciousness'. In M. Velmans and S. Schneider (eds), *The Blackwell Companion to Consciousness*. Blackwell: Malden: 313–324.

Searle, John R. (1997). *The Mystery of Consciousness*. A New York Review Book: New York.

Selfridge, Oliver. (1959). 'Pandemonium: A Paradigm for Learning'. In *Symposium on the Mechanization of Thought Processes*. H.M. Stationery Office: London: 511–526.

Seth, Anil. (2021). *Being You: A New Science of Consciousness*. Faber & Faber: London.

Simon, Jon. (2017). 'Vagueness and Zombies: Why "Phenomenally Conscious" Has No Borderline Cases'. *Philosophical Studies* 174(8): 2105–2123.

Strawson, Galen. (1992). 'The Self as Software'. *Times Literary Supplement* August 21: 5–6.