Why and How Does Consciousness Seem the Way it Seems?

Daniel C. Dennett

Are-expression of some of the troublesome features of my oft-caricatured theory of consciousness, with new emphases, brings out the strengths of the view and shows how it comports with and anticipates the recent introduction of Bayesian approaches to cognitive science.

Keywords
Bayes | Consciousness | Hume | Inversion | Qualia | Transduction

Author

Daniel C. Dennett
daniel.dennett@tufts.edu
Tufts University
Medford, MA, U.S.A.

Commentator

David Baßler
davidhbassler@gmail.com
Johannes Gutenberg-Universität
Mainz, Germany

Editors

Thomas Metzinger
metzinger@uni-mainz.de
Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt
jennifer.windt@monash.edu
Monash University
Melbourne, Australia

1 Introduction

People are often baffled by my theory of consciousness, which seems to them to be summed up neatly in the paradoxical claim that consciousness is an illusion. How could that be? Whose illusion? And would it not be a *conscious* illusion? What a hopeless view! In a better world, the principle of charity would set in and they would realise that I probably had something rather less daft in mind, but life is short, and we'll have one less difficult and counterintuitive theory to worry about if we just dismiss Dennett's as the swiftly self-refuting claim that consciousness is an illusion. Other theorists, including, notably, Nicholas Humphrey (2006, 2011), Thomas Metzinger (2003, 2009) and Jesse Prinz (2012), know better, and offer theories that share important features with mine. I toyed with the idea of trying to re-offer my theory in terms that would signal the areas of agreement and disagreement with these welcome allies, but again, life is short, and I have found that task simply too much hard work. So with apologies, I'm going to restate my position with a few new—or at least newly emphasized—wrinkles, and let them tell us where we agree and disagree.

I take one of the usefully wrong landmarks in current thinking about consciousness to be Ned Block's attempt to distinguish "phenomenal consciousness" from "access consciousness."

His view has several problems that I have pointed out before (Dennett 1994, 1995, 2005; Cohen & Dennett 2011), but my criticisms have not been sufficiently persuasive, so I am going to attempt, yet again, to show why we should abandon this distinction as scientifically insupportable and deeply misleading. My attempt should at least help put my alternative view in a better light, where it can be assayed against the views of Block and others. Here is the outline, couched in terms that will have to be clarified and adjusted as we go along:

1. There is no double transduction in the brain. (section 1)
   Therefore there is no second medium, the medium of consciousness or, as I like to call this imaginary phenomenon, the *ME*dium.
   Therefore, qualia, conceived of as states of this imaginary medium, do not exist.
2. But it seems to us that they do. (section 2)
   It seems that qualia are the source or cause of our judgments about phenomenal properties ("access consciousness"), but this is backwards. If they existed, they would have to be the *effects* of those judgments.
3. The seeming alluded to in proposition 2 is to be explained in terms of Bayesian expectations. (section 3)
4. Why do qualia seem simple and ineffable? This is an effect, a byproduct, an artifact of "access consciousness." (section 4)
5. *Whose* access? Not a witness in the Cartesian Theater (because there is no such functional place). (section 5)
   The access of other people! Our "first-person" subjectivity is shaped by the pressure of "second-persons"—interlocutors—to have practical access to what is going on in our minds.
6. A thought experiment shows how even color qualia can be understood as Bayesian projections.

2 There is no double transduction in the brain

The arrival of photons on the retina is transduced thanks to rhodopsin in the rods and cones, to yield spike trains in the optic nerve (I'm simplifying, of course). The arrival of pressure waves at the hair cells in the ear are similarly transduced into spike trains in the auditory nerve, heat and pressure are transduced into yet more spike trains by subcutaneous receptors, and the presence of complex molecules in the air we breathe into our noses is transduced by a host of different transducer molecules in the nasal epithelium. The common medium of spike trains in neuronal axons is well understood, but used to be regarded as a baffling puzzle: how could spike trains that were so alike in their physical properties and patterning underlie such "phenomenally" different phenomena as sight, hearing, touch, and smell? (see Dennett 1978, for an exposure of the puzzle.) It is still extremely tempting to imagine that vision is like television, and that those spike trains get transduced "back into subjective color and sound" and so forth, but we know better, don't we? We don't have to strike up the little band in the brain to play the music we hear in our minds, and we don't have to waft molecules through the cortex to be the grounds for our savoring the aroma of bacon or strawberries. There is no second transduction. And if there were, there would have to be a third transduction, back into spike trains, to account for our ability to judge and act on the basis of our subjective experiences. There might have been such triple transductions, and then there would have been a Cartesian Theater Deluxe, like the wonderful control room in the film Men in Black. But biology has been thrifty in us: it's all done through the medium of spike trains in neurons. (I recognize that dualists of various stripes—a genus thought extinct not so many years ago—will want to dig in their heels right here. I will ignore their howls for the time being, thinking that I can dispatch them later in the argument when I provide an answer to their implied question "What else could it be?")

So there is no *ME*dium into which spike trains are transduced. Spike trains are discriminated, elaborated, processed, reverberated, reentered, combined, compared, and contrasted—but not transduced into anything else until some of them activate effectors (neuromuscular

junctions, hormone releasers, and the like) which do the physical work of guiding the body through life. The rich and complex interplay between neurons, hundreds of neuromodulators, and hormones is now recognized, thanks to the persuasive work of Damasio and many others, as a central feature of cognition and not just bodily control, and one can speak of these interactions as transduction back and forth between different media (voltage differences and biochemical accumulations, for instance)—but none of these is the imagined *ME*dium of subjective experience.

So there just is no home in the brain for qualia as traditionally conceived. My point can be clarified by a simple comparison between two well-understood media: cinema film and digital media. First imagine showing some stone-age hunter-gatherers a movie using a portable Super-8 film projector. Amazing, they would think, but when they were then shown the frames of film up close, they would readily understand—I daresay—that this was not magic, because there were little blobs of color on each frame. (The soundtrack might still be baffling, but perhaps they would hold the film up to their ears and decide, eventually, that the sounds were just too faint for them to hear with their naked ears.) Then show them a film on a portable DVD player, and demonstrate the powers of the removable, interchangeable disks, and let them ponder the question of how such a disk managed to store all the sounds and colors they just observed on the screen. It would probably be tempting for them to declare that it *must* be magic—dualism, in other words. But with a little instruction, they could no doubt catch on to the idea that you don't have to represent color with color, sound with sound. You can *transduce* color, sound—anything, really—into a system of patterns of differences (0s and 1s, spike trains, …) and then *transduce* the elements of that system back into color and sound with playback equipment. This could lay magic to rest.

I had better make my implicit claim explicit, at the risk of insulting some readers: if you think there *has* to be a medium in the brain (or in a dualistic mind) in which subjective colors, sounds, and aromas are *rendered*, you are making the stone-ager mistake. This, I have come to believe, is the stone wall separating my view from wider acceptance. People pay attention to my arguments, and then, confronted with the prospect that qualia, as traditionally conceived, are not needed to explain their subjectivity, they just dismiss the idea as extravagant. "OF COURSE there are qualia!" This thought experiment is meant to shock them: your confidence here, I am saying, is no better grounded than the imagined confidence of the stone-agers that there just *have to be* colors and sounds on the DVD for it to convey colors and sounds to the playback machine. A failure of imagination mistaken for an insight into necessity. "But when I have a tune running through my head, it has pitch and tempo, and the timbre of the instruments is there just as if I were listening to a live performance!" Yes, and for that to be non-magically the case, there has to be a representation of the tune that progresses more or less in real time, and that specifies pitch and timbre, but that can all be accomplished without transduction, without further *rendering*, in the sequence of states of neural excitation in auditory cortex.

Vision isn't television, and audition isn't radio. We are accustomed, now, to playback devices that do transduce the signals back into the colors and sounds from which they were transduced, but we need to take advantage of our twenty-first century sophistication and recognize that the second transduction is optional! The information is in the signal, and all that information can be processed, discriminated, translated, re-coded, simplified, embellished, categorized, tagged, adjusted, and used to guide behavior without ever being transduced back into colors and sounds (or "subjective" colors and sounds).

3 It still seems that qualia exist

But it sure seems that qualia exist, in spite of the foregoing! How could they not? Aren't they needed, for instance, to be the source or cause of our judgments about them? If I have a conviction that I'm seeing an American flag after-

image (see figure 1), and note that the lowest short red stripe intersects the central cross, doesn't there have to be the red stripe I deem myself to be experiencing? Isn't the presence of that red stripe *somewhere* a necessary condition for me seeming to see a red stripe? No, and the alternative has been at least dimly understood since Hume's brilliant discussion of our experience of causation.

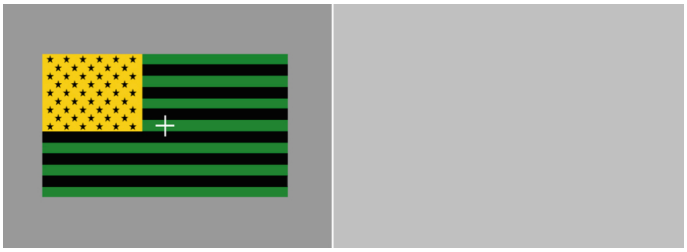

Figure 1: Inverted American Flag.

Consider what I will call Hume's Strange Inversion (cf. Dennett 2009). We think we see causation because the causation in the world directly causes us to see it—the same way round things in daylight cause us to see round things, and tigers in moonlight cause us to see tigers. When we see the thrown ball causing the window to break, the causation itself is somehow perceptible "out there." Not so, says Hume (1739, section 7 "Of the idea of necessary connexion"). What causes us to have the idea of causation is not something external but something internal. We have seen many instances of As *followed by* Bs, Hume asserts, and by a process of roughly Pavlovian conditioning (to put it anachronistically) we have been caused by this series of experiences to have in our minds a disposition, when seeing an A, to expect a B—even before the B shows up. When it does, this *felt* disposition to expect a B is mis-identified as an external, *seen* property of causation. We think we experience causation between A and B, when we are actually experiencing our internal judgment "here comes a B" and "projecting" it into the world. This is a special case of the mind's "great propensity to spread itself on external objects" (Hume 1739, I, xiv). In fact, Hume insisted, what we do is misinterpret an inner "feeling"—an anticipation—as an external property. The "customary transition" in our minds is the source of our sense of causation, a quality of "perceptions, not of objects," but we mis-attribute it to the objects, a sort of benign user-illusion, to speak anachronistically again. As Hume notes, "the contrary notion is so riveted in the mind" that it is hard to dislodge. It survives to this day in the typically unexamined assumption that all perceptual representations must be flowing inbound from outside.

> Hume wrote that the 'mind has a great propensity to spread itself on external objects' (T 1.3.14.25; SBN 167) and that we 'gild and stain' natural objects 'with the colours borrowed from internal sentiment' (EPM Appendix 1.19; SBN 294). These metaphors have invited a further one: that of 'projection' and its cognates. Though not Hume's own, the projection metaphor is now so closely associated with him, both in exegetical and non-exegetical contexts, that the phrase 'Humean projection' is something of a cliché in philosophical discourse. (Kail 2007, p. 20)

Here are a few other folk convictions that need Strange Inversions: sweetness is an "intrinsic" property of sugar and honey, which causes us to like them; observed intrinsic sexiness is what causes our lust; it was the funniness out there in the joke that caused us to laugh (Hurley et al. 2011). There is no more familiar and appealing verb than "project" to describe this effect, but of course everybody knows it is only metaphorical; colors aren't literally projected (as if from a slide projector) out onto the front surfaces of (colorless) objects, any more than the idea of causation is somehow beamed out onto the point of impact between the billiard balls. If we use the shorthand term "projection" here to try to talk, metaphorically, about the mismatch between manifest and scientific image (Sellars 1962), what is the true long story? What is literally going on in the scientific image? A large part of the answer emerges, I propose, from the predictive coding perspective. Every organism, whether a bacterium or a member of *Homo sapiens*, has a set of things in the world that matter to it and which it (therefore) needs to

discriminate and anticipate as best it can. Call this the ontology of the organism, or the organism's "Umwelt" (von Uexküll 1957). This does not yet have anything to do with consciousness but is rather an "engineering" concept, like the ontology of a bank of elevators in a skyscraper: all the kinds of things and situations the elevators need to distinguish and deal with. An animal's "Umwelt" consists in the first place of affordances (Gibson 1979), things to eat or mate with, openings to walk through or look out of, holes to hide in, things to stand on, and so forth. We may suppose that the "Umwelt" of a starfish or worm or daisy is more like the ontology of the elevator than like our manifest image. What's the difference? What makes our manifest image manifest (to us)?

4 Bayesian expectations

Here is where Bayesian expectations (see Clark 2013) could play an iterated role: our ontology (in the elevator sense) does a close-to-optimal job of representing the things in the world that matter to the behavior our brains have to control (cf. Metzinger 2003, on our world models). Hierarchical Bayesian predictions accomplish this, generating affordances galore: we expect solid objects to have backs that will come into view as we walk around them, doors to open, stairs to afford climbing, cups to hold liquid, etc. But among the things in our Umwelt that matter to our wellbeing are ourselves! We ought to have good Bayesian expectations about what we will do next, what we will think next, and what we will expect next! And we do. Here's an example:

Think of the cuteness of babies. It is not, of course, an "intrinsic" property of babies, though it seems to be. What you "project" out onto the baby is in fact your manifold of "felt" dispositions to cuddle, protect, nurture, kiss, coo over, … that little cutie-pie. It's not just that when your cuteness detector (based on facial proportions, etc.) fires, you have urges to nurture and protect; you expect to have those very urges, and that manifold of expectations just is the "projection" onto the baby of the property of cuteness. When we expect to see a baby in the crib, we also expect to "find it cute"—that is, we expect to expect to feel the urge to cuddle it and so forth. When our expectations are fulfilled, the absence of prediction error signals is interpreted as confirmation that, indeed, the thing in the world with which we are interacting has the properties we expected it to have. Without the iterated expectations, cuteness could do its work "subliminally," outside our notice; it could be part of our "elevator ontology" (the ontology that theorists need to posit to account for our various dispositions and talents) but not part of *our* ontology, the things and properties we can ostend, reflect on, report, discuss, or appeal to when explaining our own behavior (to ourselves or others). Cuteness as a property passes the Bayesian test for being an objective structural part of the world we live in (our *manifest* manifest image), and that is all that needs to happen. *Any further "projection" process would be redundant. What it is* to experience a baby as cute is to generate the series of expectations and confirmations just described. What is special about properties like sweetness and cuteness is that their perception depends on particularities of the nervous systems that have evolved to make much of them. The same is of course also true of colors. This is what is left of Locke's (and Boyle's) distinction between primary and secondary qualities.[1]

Similarly, when we feel the urge to judge something about "that red stripe" (in the American flag afterimage (see Figure 1) that hovers in our visual field, we have the temptation to insist that there is a red stripe—there has to be!—causing us to seem to see it. But however natural and human this temptation is, it must be resisted. We can be caused to seem to see something by something that shares no features with the illusory object. (Remember Ebenezer Scrooge saying to Marley's ghost: "You may be an undigested bit of beef, a blot of mustard, a crumb of cheese, a fragment of an underdone potato. There's more of gravy than of grave about you, whatever you are!" Many would insist that there has to be a ghost-shaped intermediary in the causal chain between blot of

1 The material in the previous five paragraphs is adapted from Dennett (2013).

mustard and belief in Marley, but Scrooge might be right in addressing his remark to the cause of his current condition, and be leaving nothing Marley-shaped out.) And as for the idea that without being *rendered* such contents are causally impotent, it is simply mistaken, as a thought experiment will reveal. Suppose we have a drone aircraft hunting for targets to shoot at, and suppose that the drone is equipped with a safety device that is constantly on the lookout for red crosses on buildings or vehicles—we don't want it shooting at ambulances or field hospitals! With its video eye it takes in and transduces (into digital bit streams) thirty frames a second (let's suppose) and scans each frame for a red cross (among other things). Does it have to project the frame onto a screen, transducing bit streams into colored pixels? Of course not. It can make judgments based on un-transduced information—in fact, it can't make judgments based on anything else. Similarly your brain can make judgments to the effect that there is a red stripe out there on the basis of spike train patterns in your cortex, and then act on that judgment (by causing the subject to declare "I seem to see a red stripe," or by adjusting an internal inventory of things in the neighborhood, or …). (I am deliberately using the word "judgment" for the drone's discriminations and the brain's discriminations; I have elsewhere called such items micro-takings or content-fixations. The main point of using "judgment" is to drive home the claim that these events are *not* anything like the exemplification of properties, intrinsic or otherwise. They are not qualia, in other words. Qualia—as typically conceived—would only get in the way. Don't put a weighty LED pixel screen in a drone if you want it to detect red crosses, and don't bother installing qualia in a brain if you want it to have color vision. Whatever they are, qualia are unnecessary and may be jettisoned without loss.)

So the familiar idea (familiar in the context of Block's proposed distinction between access consciousness and phenomenal consciousness) that phenomenal consciousness (= qualia) is the basis for access consciousness (= judgments about qualia, qualia-guided decisions, etc.) is backwards.[2] Once the discerning has happened in the untransduced world of spike trains, it can yield a sort of Humean projection —of a red stripe or red cross or just red, for instance—into "subjective space."

But what is this subjective space in which the projection happens? Nothing. It is a theorist's fiction. The phenomenon of "color phi" nicely illustrates the point. When shown, say, two disks displaced somewhat from each other, one sees the apparent motion of a single disk—the phi phenomenon that is the basis of animation (and motion pictures in general). If the disks are of different colors—the left disk red and the right disk green, for instance—one will see the red disk moving rightward and changing its color to green in mid-trajectory. How did the brain "know" to move the disk rightward and switch colors before having access to the green disk at its location? It couldn't (supposing precognition to be ruled out). But it could have Bayesian expectations of continuous motion from place to place that provoke a (retrospective) expectation of the intermediate content, and this expectation encounters no disconfirmation (if the timing is right), which suffices to establish in reality the illusory sequence in the subject's manifest image. So the visual system's *access* to the information about the green disk is causally prior to the "*phenomenal*" motion and color change. Here is a diagram of color phi

2 I once had an occasion to point out this prospect to Block. He had just participated in a laterality test, to see how strongly lateralized for language his brain was. There are two oft-used ways of testing this: with dichotic headphones, which send different words into each ear, where the subject is asked to identify the word heard (typically you only hear one of them!). A second, visual test involves looking at a center target on a screen and having a word or non-word (e.g., "flum" or "janglet") flashed briefly in either the left or right visual field. The subject presses the word button or the non-word button and latencies and errors are recorded. If you are strongly lateralized left (your left hemisphere is strongly dominant for language and does most of the work of language processing), you are faster and more accurate on words and nonwords flashed to the right hemifield. Ned had taken the visual test, and I asked him what he had learned. He was, he said, strongly lateralized left for language, like most people, and he added "the words flashed on the left actually seemed blurry!" I asked him whether the words seemed blurry because he noted the difficulty he was having with them, or whether he had the difficulty because the words were blurry. He acknowledged that he had no introspective way of distinguishing these two hypotheses. Supposing that Block doesn't have some remarkable problem with his eyes, in which the left half of each lens is occluded or misshapen, producing a blur on the left side of his retinas, it is highly likely that the blurriness he seemed to experience was an effect of his felt difficulty in responding, not the cause of this difficulty.

from Consciousness Explained (and Dennett & Kinsbourne 1992):

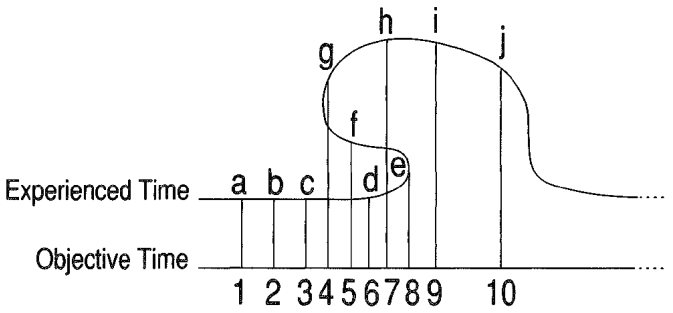

Figure 2: Superimposition of subjective and objective sequences.

In order to explain "temporal anomalies" of conscious experience, we need to appreciate that not only do we not have to represent red with something red, and round with something round; we don't have to represent time with time. Recall my example "Tom arrived at the party after Bill did." When you hear the sentence you learn of Tom's arrival before you learn of Bill's, but what you learn is that Bill arrived earlier. No revolution in physics or metaphysics is needed to account for this simple distinction between the temporal properties of a representation and the temporal properties represented thereby. It is quite possible (in color phi, for instance) for the brain to discern (in objective time) first one red circle (cat time 3) and then a green circle (fat time 5) displaced to the right, and then to (mis-)represent an intermediate red-turning-green circle (eat time 8) yielding the subjective judgment of apparent motion with temporally intermediate color change. Here our Bayesian probabilistic anticipator is caught in the act, jumping to the most likely conclusion in the absence of any evidence. Experienced or subjective time doesn't line up with objective time, and it doesn't have to. The important point to remember from the diagram is that the subjective time sequence is NOT like a bit of kinked film that then has to be run through a projector somewhere so that c is followed by e is followed by f in real time. It is just a theorist's diagram of how subjective time can relate to objective time. Subjective time is not a further real component of the causal picture. No *rendering* is necessary, the judgment is already in, and doesn't have to be re-presented for another act of judging (in the Cartesian Theater).

The temptation to think otherwise may run deep, but it is fairly readily exposed. Consider fiction. Sherlock Holmes and Watson seem real when one is reading a Conan Doyle mystery —as real as Disraeli or Churchill in a biography. When Sherlock seems real, does this require him and his world to be rendered somewhere, in—let's call it—*fictoplasm?* No. There is no need for a medium of fictoplasm to render fiction effective, and there is no need for a mysterious medium, material or immaterial, to render subjective experience effective. No doubt the temptation to posit the existence of fictoplasm derives from our human habit, when reading, of adding details in imagination that aren't strictly in the book. Then, for instance, when we see a film of the novel, we can truly say "That's not how I imagined Holmes when I read the book."

Isn't such rendering in imagination while reading a novel a case of *transduction* of content from one medium (written words as seen on the page) into another (imagined events as seen and heard in the mind's eye and ear)? No, this is not transduction; it is, more properly, a variety of *translation, effortlessly expanding the content thanks to the built-in Bayesian prediction mechanisms.* We could construct, for instance, a digital device that takes problems in plane geometry presented in writing ("From Euclidean axioms prove the Pythagorean Theorem.") and solves them through a process that involves making Euclidean constructions, with all the sides and angles properly represented and labeled, and utilizing them in the proof. The whole process from receipt of the problem to delivery of the called-for proof (complete with printed-out diagrams if you like), is conducted in a single medium of digital bit strings, with no transduction until the printer or screen is turned on to render the answer. (A more detailed description of this kind of transformative process without transduction is found in my discussion (1991) of how the robot Shakey discriminated boxes from pyramids.)

Consider Figure 2 above. Does the access/phenomenal consciousness distinction get depicted therein? If so, access consciousness should be identified with the objective time line, and phenomenal consciousness (if it were something real in addition to access consciousness) would be depicted in the line that doubles back in time. The content feature that creates the kink is an effect of a judgment or discernment that came later in objective time than the discernment of the green circle at time 5. It is because the brain already had access to red circle, then green circle that it generated a representation (but not a *rendering*) of the in-between red-turning-green circle as an elaborative effect.[3]

5 Why do qualia seem so simple and ineffable?

Qualia seem atomic to introspection, unanalyzable simples—the smell of violets, the shade of blue, the sound of an oboe—but this is clearly an effect of something like the resolution of our discernment machinery.

If our vision were as poorly spatially resolved as our olfaction, when a bird flew by, the sky would suddenly "go all birdish,"—that peculiar, indescribable birdishness that one would experience in the visual presence of birds. And this resolution is variable: music lovers and wine enthusiasts and others can train up their ear and their palate and come to distinguish, introspectively, the combining elements of what used to seem atomic and unanalyzable. David Huron (2006), has done some ingenious work teasing out and explaining the combinations of neuroarchitectonic properties that explain the otherwise ineffable characteristic qualia of scale tones (the way *do* sounds different from *re* and *mi* and *so*). It turns out that these "qualia" are actually highly structured properties of neural representations. The explanation, needless to say, is ultimately in the medium of spike trains.

But why should the resolution (if that is the right term) be so low? Why should our brains ignore so much detail in the representations to which "we" have "access"? Minsky (1985), Dennett (1991), Norretranders (1999), Metzinger (2003), and others have said that it is the brain's own access to its own complex internal activities that accounts for the simplicity. This is the brain's effective user-illusion for itself, in much the way the desktop with its icons and various metaphors (click and drag, highlighted targets, etc.) is an elegantly designed user-illusion for laypeople who don't need to know how their computers work.

The brain does not have a single internal witness or homunculus, but it does need something like a lingua franca to get the different and semi-independent subsystems to communicate with each other. (For instance, in the Global Neuronal Workplace model[4] of Dehaene et al. (2006), and others, one should not take it for granted that the *local* meanings of spike train patterns—in the dorsal vision stream, say, or the olfactory bulb—are readily "understood" by all the elements to which some of these signals are broadcast.) I think there is bound to be some important truth in that theme, but it is only part of the story.

6 Whose access?

I think the more interesting suggestion is that the effective "we" when we talk about what "we" have access to, is, indeed, *we*—not just *I*, but *you and me.* It is, more particularly, *your* access to *my* mind that simplifies the information that *we* have access to!

The linguist Stephen Levinson (2006) has studied the remarkable language, Yélî Dnye, of the three thousand or so inhabitants of Rossel Island in the South Pacific—to the north of Papua New Guinea. It is a completely isolated language, unlike any other in the world in many regards. In particular, it is hideously complex, with:

> the largest phoneme inventory (ninety distinct segments) in the Pacific, and many

3 Thanks to David Gottlieb for drawing my attention to this way of looking at access consciousness.

4 Isn't the Global Neuronal Workplace the derided Cartesian Theater after all? No, because what goes on there is not transduction-and-rendering, but informational integration: the coalition and consilience of competing elements. There is no transduction threshold that determines the time-of-entry "into consciousness", and none of the multiple drafts competing in it are singled out as being conscious except retrospectively. This is the point of my admonition always to ask the Hard Question: "And then what happens?" (Dennett 1991, p. 225)

> sounds (such as doubly articulated labial coronal stops) that are either unique or rare in the languages of the world. Among the fifty-six consonants are many multiply articulated segments: e.g., /ʇpn̪m/ is a single segment made by simultaneously putting the tongue behind the alveolar ridge, trilling the lips , and snorting air through the nose. [...] Once the learner is past the sound hurdle, he or she faces another formidable obstacle. The language has an extremely complex system of verb inflection (with thousands of distinct inflectional forms). [...] In addition, substitute forms are used where the subject has been mentioned before, is close or visible, is in motion, or where the sentence is counterfactual or negative, thus providing well over a thousand possibilities [...]. (Levinson 2006, p. 20)

Levinson reports, not surprisingly, that "[h]ardly any mature individuals (such as non-native spouses) who have immigrated into the island community ever learn to speak the language, and children of expatriate Rossels do not fully acquire it from their parents alone." His explanation is speculative, but plausible: a language, left to itself for centuries, will grow ever more complex, like an unpruned bush, simply because it can. The extreme isolation of Rossel Island over the centuries (for various geographic reasons) means that the language has hardly ever been confronted with non-native speakers of another language with whom communication is imperative, for one reason or another. The need for communication soon generates a small cadre of bi-lingual interpreters, and maybe also a pidgin (and maybe later a creole), and all of these alien interfaces work to simplify a language. The least learnable, most baroque (in the sense of exceeding the functional) features of the language are dropped under this pressure. We can see it happening with English today, with simplified dialects such as Emblish (as spoken at the European Molecular Biology Laboratory in Heidelberg) arising naturally and imperceptibly.

I would like to speculate that a similar process of gradual but incessant simplification has shaped the language we have available to explain and describe our minds to each other. Wittgenstein's famous claim about the impossibility of a private language has not weathered the storms of controversy particularly well, but there are neighboring claims—empirical claims—that deserve consideration. Many years ago, Nicholas Humphrey (1987) made the point that has begun to attract adherents today:

> While it is of no interest to a person to have the same kind of kidney as another person, it is of interest to him to have the same kind of mind: otherwise as a natural psychologist he'd be in trouble. Kidney transplants occur very rarely in nature, but something very much like mind-transplants occur all the time [...]. [So] we can assume that throughout a long history of evolution all sorts of different ways of describing the brain's activity have been experimented with but only those most suited to doing psychology have been preserved. Thus the particular picture of our inner selves that human beings do in fact now have—the picture we know as 'us', and cannot imagine being of any different kind—is neither a necessary description nor is it any old description of the brain: it is the one that has proved most suited to our needs as social beings. That is why it works. Not only can we count on other people's brains being very much like ours, we can count on the picture we each have of what it's like to have a brain being tailor-made to explain the way that other people actually behave. Consciousness is a socio-biological product—in the best sense of socio and biological. (p. 18)

Chris Frith, for instance, has recently taken up the theme (in conversation) that consciousness has some features, because everything in consciousness has to be couched in terms that can be communicated to other people readily.

The ineffability barrier we all experience when trying to tell others what it is like to be

us on particular occasions is highly variable, not just between individuals, but over time within a single individual, as a result of formal or informal training. It plays a dynamic role in shaping the contents of our consciousness over time.[5] (This would be true only for human consciousness, obviously.)

7 A thought experiment: Mr. Capgras

Finally, it might seem that whereas some subjective properties—cute, sweet, funny, sexy, the characteristic sounds of scale tones—might be accounted for in terms of Bayesian expectations about how one will be disposed to behave in their presence, the very simplicity of colors must block any attempt to treat them in a similar fashion. There is no way one expects to behave in the presence of navy blue, or pale yellow, or lime green. So it may seem, but this is itself an artifact of our penchant for thinking—as Hume famously did—of colors as simples. Hume was discountenanced by the notorious missing shade of blue, and found it ideologically inconvenient to suppose, as we now know, that color experience is in fact highly complex and compositional, and deeply anchored in dispositions of our perceptual systems.[6] Moreover, color experiences are no more atomic than scale tone experiences, and give rise to all manner of expectations, which tend to go unnoticed, but can be thrown into sharp focus by a thought experiment: my fantasy about poor Mr. Clapgras, the man who wakes up to find all his emotional dispositions with regard to colors inverted while leaving intact his cognitive habits and powers (see Dennett 2005, pp. 91–102, for a more detailed account, with objections considered and rebutted). Ex hypothesi, Mr. Clapgras identifies colors and sorts colors correctly (he does not suffer from the well-studied conditions color anomia, or cerebral achromatopsia), but he finds the world disgusting, unbearable. Food looks just terrible to him now, and he has to eat blindfolded, since his emotional responses to all colors have shifted 180 degrees around the color circle (Grush this collection). He calls shocking pink "shocking pink" but marvels at the inappropriateness of its name. The only way we can explain his distress is by observing that he notices that something is wrong—which has to mean he was expecting something else. He is surprised that breaking a fresh egg into a frying pan on a sunny morning doesn't bring a smile to his face, that a glimpse of his obnoxious neighbor's lime green convertible doesn't irritate him the way it used to do, that he feels no stirring of childhood patriotism when he sees the red white and blue waving in the breeze. Like the sufferers of Capgras delusion, poor Mr Clapgras senses a disturbance: something is very wrong, but it isn't the evaporation of intrinsic internal properties.

8 Conclusion

The considerations I have raised in this essay are not new, but perhaps bringing them together as I have done will help show that a counter-intuitive theory like mine still has an advantage over some of the fantasies in which philosophers have recently indulged. It may well be, as Paul Bloom (2004) has suggested, that we are all "natural born dualists," but just as eyeglasses can correct for myopia, natural-born or not, so science can correct for this innate cognitive disability. Intuitions to the contrary are important data, but should not be taken to indicate a limitation of science, as some have thought. In fact, if the best scientific theory of consciousness turns out not to be deeply counterintuitive at first, among the data it will have had to explain is why it took us so long to arrive at it.

Acknowledgements

Thanks to Michael Cohen, and to Andy Clark, and the rest of Dmitry Volkoff's Greenland Consciousness and Free Will workshop (June 10-17, 2014), for editorial advice on this essay.

5 Note that I am not saying that our day-to-day consciousness wouldn't occur in the absence of human company, but an implication of my speculation is that a Robinson Crusoe human, somehow raised from birth without human contact, would have subjectivity more inaccessible to us—once we discovered him and attempted to communicate with him—than the speech acts of the Rossel Islanders.

6 In Cohen & Dennett, we point out that limbic or emotional responses to colors have to count as instances of "access" to color-representing states "however coarse-grained or incomplete, because such a reaction can obviously affect decision making or motivation" (2011, p. 5).

References

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human.* New York, NY: Basic Books.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36* (3), 181-253. 10.1017/S0140525X12000477

Cohen, M. A. & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15* (8), 358-365. 10.1016/j.tics.2011.06.008

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10* (5), 204-211. 10.1016/j.tics.2006.03.007

Dennett, D. C. (1978). "What's the difference: Some riddles," (commentary on Puccetti and Dykes). *Behavioral and Brain Sciences*, *1* (3), 351-351.

——— (1991). *Consciousness explained.* Boston, MA: Little, Brown and Company.

——— (1994). Get real. *Philosophical Topics*, *22* (1-2), 505-568.

——— (1995). "The path not taken," commentary on Ned Block, "On confusion about a function of consciousness". *Behavioral and Brain Sciences*, *18* (2), 252-253.

——— (2005). *Sweet dreams: Philosophical obstacles to a science of consciousness.* Cambridge, MA: MIT Press.

——— (2009). Darwin's 'strange inversion of reasoning'. *Proceedings of the National Academy of Sciences*, *106* (1), 10061-10065. 10.1073/pnas.0904433106

——— (2013). Expecting ourselves to expect: The Bayesian brain as a projector (commentary on Clark, 2013). *Behavioral and Brain Sciences*, *36* (3), 209-210. 10.1017/S0140525X12002208

Dennett, D. C. & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences*, *15* (2), 183-247. 10.1017/S0140525X00068229

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton Mifflin.

Grush, R., Jaswal, L., Knoepfler, J. & Brovold, A. (2015). Visual Adaptation to a Remapped Spectrum: Lessons for Enactive Theories of Color Perception and Constancy, the Effect of Color on Aesthetic Judgments, and the Memory Color Effect. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.

Hume, D. (1739). *Treatise of human nature.* London, UK: John Noon.

Humphrey, N. (1987). *"The uses of consciousness" The 57th James Arthur Lecture.* New York, NY: American Museum of Natural History.

——— (2006). *Seeing red: A study in consciousness.* Cambridge, MA: Harvard University Press.

——— (2011). *Soul dust: The magic of consciousness.* Princeton, NJ: Princeton Universiry Press.

Hurley, M., Dennett, D. C. & Adams , jr. , R. B. (2011). *Inside jokes: Using humor to reverse-engineer the mind.* Cambridge, MA: MIT Press.

Huron, D. (2006). *Sweet anticipation; Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Kail, P. J. E. (2007). *Projection and realism in Hume's philosophy.* Oxford, UK: Oxford University Press.

Levinson, S. C. (2006). Introduction. In S. C. Levinson & P. Jaisson (Eds.) *Evolution and culture* (pp. 1-42). Cambridge, MA: MIT Press.

Metzinger, T. (2003). *Being no one. The self-model theory of subjectivity.* Cambridge, MA: MIT Press.

——— (2009). *The ego tunnel. The science of the mind and the myth of the self.* New York, NY: Basic Books.

Minsky, M. (1985). *The society of minds.* New York, NY: Simon & Schuster.

Norretranders, T. (1999). *The user illusion: Cutting consciousness down to size.* London, UK: Penguin Press Science.

Prinz, J. (2012). *The conscious brain.* Oxford, UK: Oxford University Press.

Sellars, W. (1962). Philosophy and the scientific image of man. In R. Colodny (Ed.) *Frontiers of science and philosophy* (pp. 35-78). Pittsburgh, PA: University of Pittsburgh Press.

von Uexküll, J. (1957). A stroll through the worlds of animals and men: A picture book of invisible worlds. In C. H. Schiller (Ed.) *Instinctive behavior: The development of a modern concept.* New York, NY: International Universities Press.