Michael Tye argues for an externalist version of representationism concerning the phenomenology of perception. Representationism on one version says that what it is for a percep to have a certain phenomenal character is for it to have a certain representational content. Tye wants to stay neutral on which form of representationism he is defending, but the kind of representationism he has advocated and I have criticized is externalist and reductionist. What it is to have an experience as of redness is to be in a state of nonconceptually representing an external world property redness. (Tye adds a functional condition to distinguish conscious from unconscious perception, and there are restrictions on the kind of content involved, but I won’t discuss those aspects of the view.) Tye’s view is distinct from what one might call a phenomenology-first view in which the property represented is phenomenal or in which there is an irreducible phenomenal mode of presentation. It would be tempting to describe Tye’s view as a representation-first view since he thinks there is nothing more to phenomenal character than representational content, but since Tye’s view is an identity theory (phenomenal character and the representational property are one and the same), it is better termed what Adam Pautz (2008, 263) calls a “no-priority” view.

What is front and center in Tye’s chapter is not representationism per se but the putative historical dimension of externalist representation: the content of a representation that a subject has at noon depends constitutively (and not just causally) on what happened at previous times and even on the evolutionary history of the species. Although the official topic is representationism, the issues raised are also highly relevant to functionalism.

The argument proceeds by comparing three different science fiction examples:

*Group-satellite brain:* This case is descended from Hilary Putnam’s (1967) example of a swarm of bees that duplicates human functional organization. A large number of people communicating by satellite play the role of an external brain for a robot body. I used the nation of China as an example because of its large population (Block 1978), and that led to Searle’s Chinese Room argument two years later. (Searle told me at the time that he had read my paper.) The distributed brain and robot body together
duplicate human functional organization. Importantly, the group-satellite brain need not be based on human neuroscience. (Actually, there is a psychofunctional simulation that is based on human psychology and a physiofunctional version based on human neuroscience, but Tye does not seem to have either of those in mind.) Rather, it is devised on the basis of aspects of human functional organization that are available to common sense, what I once called “Analytic Functionalism”—a variant of what David Chalmers (1996) calls Type A materialism (Block 1978).

**Elementary-particle simulators:** In a part of the universe where matter is infinitely divisible, intelligent creatures use their spaceships to create (out of their matter) particles with the chemical and physical characteristics of our elementary particles. Their ships simulate the three kinds of elementary particles (updating from the physics of 1978): (1) leptons of various sorts (e.g., electrons), (2) quarks of various sorts, and (3) bosons (e.g., photons). They reproduce the various energy transactions and interactions of these particles. For example, a quark with a certain color together with an antiquark with the anticolor of that color form a color-neutral meson, and likewise for the simulated matter. The science fiction example continues: you discover this simulated matter, mine it, grow food on it, make breathable air, and so on. Over time you come to be made mainly of this simulated matter (Block 1978).

**Silicon replacement:** This example may derive from an example used by John Hauge-land (1980), but the best version and the one that Tye uses is due to David Chalmers (1995). There are two stages.

- Stage 1: Cell bodies are replaced one by one by silicon chips that process inputs from dendrites and outputs to axons just as real cell bodies do. The chips have to have “transducers” that transform the chemical inputs and outputs of these biological input and output devices. Stage 1 is a hybrid, combining elements of our electro-chemical mechanisms with digital chips.
- Stage 2: Once groups of neurons are so replaced, the “awkward axons and dendrites that mediate the connection between the chips” are simply replaced with standard digital connections and by the time all these replacements have happened, there are “no biochemical mechanisms playing an essential role.” (Chalmers 1995, 314)

On the group-satellite brain, Tye agrees with me that it has no phenomenology. According to me, that is because it lacks the biological mechanisms that underlie phenomenology, but according to him, that is because of its history: it is not naturally evolved. He also agrees with me that when we come to be made out of elementary-particle simulators, our phenomenology does not change, but he thinks that the reason is not the one I gave—that biological mechanisms are preserved—but rather that our evolutionary history is preserved with the change in our matter.

Tye’s claims about the importance of history are based on the silicon replacement case. Tye and Chalmers think the silicon replacement does not change phenomenology.
On my view, human phenomenology probably depends on our biological machinery. Tye thinks that silicon replacement does not change phenomenology because evolutionary history is (allegedly) preserved. Chalmers thinks that there is a nomological correlation between functional organization and phenomenology such that certain functional states are nomologically sufficient for consciousness. That disagreement with me on one side and both Tye and Chalmers on the other is the main topic of this response. I start with the silicon replacement and what Chalmers says about it.

The gradualness of the silicon replacement is crucial to Chalmers’s argument. The starting point is Conscious Dave with his rich conscious experience, and the ending point is the silicon-brained being—following Chalmers, we can call him Robot. A sample intermediate case—let us suppose somewhere in the middle of the transformation—is Joe. Following Chalmers and Tye, let us start by supposing that Robot is a nonconscious zombie. Chalmers and Tye ultimately reject the claim that Robot is non-conscious, but their form of argument is to start by supposing it is true and using that to generate unacceptable consequences.

Consciousness disappears during the gradual replacement of Conscious Dave’s neurons, dendrites, and axons by silicon simulators. According to Chalmers, the two possibilities for intermediate cases are Suddenly Disappearing Qualia and Fading Qualia. On the first option, the replacement of a single neuron destroys consciousness altogether. Chalmers gives a broadly empirical argument that this option would require an unprecedented brute discontinuity in laws of nature. My own view is that there is a closely related empirical reason that is more powerful: it is fundamental to the way the brain works that everything is accomplished by neural circuits, not individual neurons. There has never been an accepted neural model of any brain process (as far as I know) in which the destruction of a single neuron has massive effects. In sum, I agree with Chalmers that this option has very little empirical plausibility.

The second option is that the intermediate case, Joe, has faint versions of all Conscious Dave’s experiences; for example, he sees tepid pink where Dave sees bright red. But, of course, it is built in to the example that what he says and judges about his experiences is the same as what Dave says and judges about his quite different experiences, or at least that the words that Joe utters are the same as the words that Dave utters. (I ignore the possibility that Joe’s words have different meanings from Dave’s.)

If consciousness is an all-or-nothing thing, a gradual replacement of its neural basis will not result in a gradual decrease in consciousness, so the fading qualia thought experiment may be ruled out from the start. Even if consciousness is in some sense gradual, it is not clear that losing it gradually can be coherently described. A perception of tepid pink can be just as conscious as a perception of bright red. And as came out in a discussion of this reply at MIT, it is not clear that any other picture of the gradual loss can work. I’ve heard loss of consciousness with general anesthetic described as the shrinking of the visual field to a point. But if Joe’s visual field is shrunken, how could
he process visual information in an equivalent manner to Conscious Dave? Similar points could be made about Joe feeling increasingly drowsy as more cells are replaced by chips. The problem is that loss of consciousness has behavioral ramifications, but behavioral differences are ruled out ex hypothesi.

However, for the sake of argument, I will buy into the idea that it makes sense to describe Joe as less conscious than Conscious Dave. So what Joe says and judges is systematically wrong about his experience. And that, according to Chalmers, is empirically implausible:

There is a significant implausibility here. This is a being whose rational processes are functioning and who is in fact conscious, but who is completely wrong about his own conscious experiences. Perhaps in the extreme case, when all is dark inside, it is reasonable to suppose that a system could be so misguided in its claims and judgments—after all, in a sense there is nobody in there to be wrong. But in the intermediate case, this is much less plausible. In every case with which we are familiar, conscious beings are generally capable of forming accurate judgments about their experience, in the absence of distraction and irrationality. For a sentient, rational being that is suffering from no functional pathology to be so systematically out of touch with its experiences would imply a strong dissociation between consciousness and cognition.

We have little reason to believe that consciousness is such an ill-behaved phenomenon, and good reason to believe otherwise.... Unless we are prepared to accept this massive dissociation between consciousness and cognition, the original system must have been conscious after all. (1995, 316; my italics)

One could raise the question of whether the silicon simulation is really possible. There are many mechanisms of neural information transfer that on the face of it may be difficult or impossible to simulate in real time in a small space. Neurons affect other neurons in part by many types of complex mechanisms (for example, slow profusion of neurotransmitters into extracellular fluid). And some transfers of information work via direct connections between neurons (“gap junctions”) through which many types of molecules can flow from one neuron to another—rather than via a synapse. But I put these issues aside for the moment and assume that the scenario that Chalmers describes is indeed possible.

What to think about Joe is the crux of the issue. Chalmers’s argument again is broadly empirical. He says of Joe that his “rational processes are functioning,” and speaking of Joe, he says, “For a sentient, rational being that is suffering from no functional pathology to be so systematically out of touch with its experiences would imply a strong dissociation between consciousness and cognition.”

What does it mean to say that Joe suffers from no functional pathology? There is an important ambiguity in this notion. One of the normal functional roles of experiences as of pink is to enable introspective judgments that one is seeing something pink, but Joe does not instantiate this normal introspective relation. Instead Joe makes false judgments to the effect that he is seeing something bright red. And we may
suppose that Joe insistently denies seeing something pink. Joe has phenomenal states that—pathologically—do not have the kinds of effects on belief as in a functionally normal person. Joe’s introspection is systematically unreliable—so why isn’t that functional pathology?

Of course, there is another notion of “functional pathology” that does not take into account the phenomenal character of the functional state. Joe’s functional organization is isomorphic to Conscious Dave’s functional organization. Joe is functionally normal in that purely causal sense, a sense that ignores phenomenal character.

One rationale for Chalmers’s notion of “functional pathology” is that it might seem that there could be no functional pathology without behavioral pathology. And Joe’s behavior, it seems, is perfectly normal. We may reasonably suppose that he doesn’t run afoul of the law or end up in the hospital or lose his job. But these behavioral tests are not the only tests of functional pathology. It is a kind of functional pathology to howl as if one is in excruciating pain, saying that one is in excruciating pain, when one is experiencing the pain equivalent of tepid pink.

Here is the crux: there is one functional regularity that all should agree is preserved in Joe and another functional regularity that on the biology-centered view that I hold is not preserved. I cannot see a theory-neutral rationale for preferring one to the other. Chalmers’s rationale for focusing on the phenomenology-neutral notion would seem to be functionalism itself. That is, the rationale would be that the phenomenal character of a state is assumed to be determined by its function. But to appeal to functionalism in a defense of functionalism is question-begging.

However, instead of the blanket term “functional pathology,” Chalmers might simply have given a list of mental states and conditions that in cases with which we are familiar make introspection unreliable, such as distraction and self-deception. And none of those circumstances are in play in Joe, the intermediate case. We could regard “functional pathology” as an abbreviation of “distraction, self-deception, and other conditions that in cases with which we are familiar make introspection unreliable.”

But a list-like notion of “functional pathology”—based on armchair considerations—is greatly inferior to one based on mechanisms, and to find the mechanisms you have to leave the armchair. Many cases of failure of reliability of a person’s judgments about his or her own experience have nothing to do with “cases with which we are familiar” in daily life. Perhaps the most dramatic case is anosognosia, in which patients systematically deny a deficit—even while complaining about other deficits (Marcel 2004). A particularly interesting case is anosognosia for hemiplegia (Block 2011; Fotopoulou et al. 2010; Marcel 2004). Hemiplegia is paralysis of one side of the body. In anosognosia for hemiplegia, subjects who are told to raise their arm and fail to do it seem incapable of appreciating that they are experiencing the arm not moving. In some cases of anosognosia, denial of a deficit may be due to hallucination, but as Fotopoulou and colleagues argue, that is probably not the case for this condition. So the
subject may be experiencing the arm not moving while simultaneously claiming it is moving.

I have seen quite a few videos of these patients, and no patient I have seen has been asked while trying to move his arm, “Are you having the experience as of moving your arm?” The patients and doctors who question them do not put their points in the ways philosophers might want them to. But patients do say they are moving their arms when they can see perfectly well that the arms are not moving.

Anosognosics for hemiplegia are clear cases of functional pathology in Chalmers’s sense as well as mine: they fall down because they don’t know that they are paralyzed on one side. But they are a useful corrective to the idea that there cannot be a massive disconnect between phenomenology and belief about phenomenology.

No good notion of functional pathology can be framed without considering mechanisms by which experience produces judgments and behavior, and to find them we must leave the armchair. Thus, I think that the methodology used by both Chalmers and Tye is flawed. I now approach this same point from a slightly different direction. It is useful to proceed using the example of the quantum approach to consciousness. I do not take this approach seriously, nor do very many neuroscientists take it seriously, but it is useful for illustrating the point.

Stuart Hameroff and Roger Penrose have proposed that consciousness depends on quantum processes inside tiny microtubules that are part of the skeleton of cells and are located inside neuronal cell bodies, axons, and dendrites (Hameroff and Penrose 2014). Microtubules are part of the cytoskeleton of the cell that maintains the shape of the cell. Whatever one thinks of this theory, it is not refuted by Chalmers’s thought experiment. Robot, as Chalmers describes it, would lack consciousness according to the Hameroff-Penrose account because of the lack of microtubules in the silicon chips. (I am assuming that they hold that quantum processes of the sort in microtubules are necessary for consciousness, so a silicon device that simulates us need not have such processes.) And intermediate cases would also be deficient in microtubules though not totally devoid of them. So at some point in the progression from Conscious Dave to zombie Robot, the deficit in microtubules could be expected to result in a deficit in consciousness even though that intermediate case (Joe) is guaranteed by the terms of the thought experiment to walk and talk and would be, more generally, functionally just like Conscious Dave.

Of course, Hameroff and Penrose may take the line that no mechanism that lacks microtubules can possibly duplicate the functional organization of a conscious human. I am tempted to say the same about the biological mechanisms I envision as the ground of consciousness. But let us put that issue aside and assume that the zombie Robot is possible.

I doubt that even with this concession, Hameroff and Penrose would regard their view as refuted by Chalmers’s argument—nor should they. What they should say is that,
in “every case with which we are familiar,” our conscious utterances and consciously controlled behavior are caused by microtubule activity. So how can the postulation of a different mechanism by which a robot’s utterances and behavior are caused by something else refute the microtubule account of us? Our introspections are accurate. Joe, lacking as he does some of the quantum mechanisms that ground consciousness and that are necessary for consciousness, has inaccurate introspections. Whatever oddity there is in the consideration of Joe derives from constructing a case in which what third parties can observe stays the same, whereas the mental states that the observables tell us about change.

Suppose I text myself “Just ran out of milk” to remind myself to buy milk. The text is evidence that I used to have milk. But if the world had come into existence a fraction of a second ago complete with all the evidence of the past (as Russell once considered), then all evidence of the past—including my text—is misleading. Cases with which we are familiar provide no precedent for such massive unreliability, but the absurdity comes from the setup, not any impossibility of the claim that the world came into existence a fraction of a second ago.

As I said, I do not accept the microtubule account, so how can I use it to defend my account? My account is much vaguer—that there is something about our biological makeup—perhaps its electrochemical character—that underlies conscious phenomenology. So the objection to Chalmers and Tye based on my account would be the same as the one given for the microtubule theory—except in its appeal to something so specific as quantum activity in microtubules.

Tye’s version of the argument is importantly different from Chalmers’s. He does not appeal to empirical plausibility—to the presumed reliability of our cognitive appreciation of our own phenomenology in those who do not suffer from functional pathology—but rather to “plausibility” in a sense he does not explain but seems to be some sort of armchair intuition. For example, he says, of the idea that Robot’s phenomenal beliefs are all wrong, that it is very implausible. It requires us to accept that the being at the end of the replacement process is radically mistaken about his own phenomenal life even though he is fully rational. This is difficult to swallow. A more reasonable hypothesis is that rational conscious beings are not so mistaken.

Although Tye and Chalmers both appeal to the rationality of the subject of the thought experiment, there is a difference: At every crucial point, Tye appeals to “plausibility.” I count ten occurrences of variants of the word, not counting other words with the same use like “swallow” or considerations of what it is “reasonable to say.” This argument has the same weaknesses as Chalmers’s argument plus the additional weakness of supposing the argument from familiar cases is all a matter of plausibility rather than an empirical consideration of the nature of consciousness.
Although I have emphasized that Tye is giving an armchair argument, whereas Chalmers (1995) appeals to the broadly empirical, Chalmers’s argument has strong a priori elements. See especially the very interesting online dialogue on Consciousness Online involving comments on a paper by Miguel Sebastian (2014). In that thread, Adam Pautz notes that since on Chalmers’s view there are possible worlds in which there are fading qualia, it has to be a contingent truth that in our world there can be no fading qualia. But if the arguments for the impossibility of fading qualia in the actual world are all a priori, then the claim that there can be no fading qualia in the actual world would be a strange case of the contingent a priori. Chalmers’s reply does not reject the contingent a priori. Instead, Chalmers appeals to the idea that nondemonstrative reasoning such as inductive reasoning can support a conclusion even though there are counterinductive worlds in which that reasoning would go wrong. In this discussion, Chalmers’s focus is on plausibility rather than anything even broadly empirical.

Tye introduces one further wrinkle. He supposes that the word “pain” as used by the individual in which the chips are implanted expresses a “concept...that rigidly picks out a state whose essence is its phenomenal character.” After the individual’s neurons are replaced by silicon chips, the word as used by the zombie that results from the replacement rigidly picks out a nonphenomenal state. So a new concept has been introduced. He then argues that the introduction of this new concept would be puzzlingly unlike all other cases of concept introduction, concluding,

It seems to me, then, that the most reasonable thing to say is that there is just a single concept expressed by “pain” and that during the replacement process the beliefs do not change and neither does the phenomenology. So if we wish to respect our initial intuitions and we also wish to avoid getting embroiled in puzzles and problems that arise once we take the view that the phenomenology changes with the gradual silicon chip replacement, we should accept that, notwithstanding the absence of sameness in electrochemical mechanisms, there is sameness in phenomenology.

Suppose that Robot does not lack consciousness altogether but simply has a different kind of consciousness from us. And suppose further that Joe has a kind of consciousness that bears some similarity to ours and some to Robot’s. In that case, the phenomenal pain concept used by Robot and Joe would be different from ours. (I am thinking of phenomenal concepts as encompassing some sort of “sample” of a phenomenal state (Balog 2009; Block 2006; Papineau 2002). We can use this account to explain why Conscious Dave’s, Joe’s, and Robot’s phenomenal concepts differ from one another. And if Robot completely lacks consciousness as the argument supposes, the “sample” would be a sample of nothing, making Robot’s “phenomenal concept,” if it can be called that, very different from ours.

In short, Robot’s and Joe’s concepts would have samples that are unlike ours. That this would be unlike other cases of the introduction of a phenomenal concept seems
to me to count against it not at all. And the same applies to the limiting case in which
the sample is a nonphenomenal state.

Similar issues are at play in Tye’s and my divergent treatments of the elementary-
particle simulators. I argue that since the biological mechanisms are preserved, so is
conscioussness. Tye objects on two grounds: first, counterfactuals are true of us after
the replacement that are not true before it. If the elementary-particle people decide to
abandon the project after the replacement, we die, whereas abandoning the project
before the replacement won’t affect us. I say that this is on a par with hardware failure
in any system. If you build a Mac laptop out of this material, it will have different fail-
ure characteristics than if you build it out of ordinary matter. Whether this difference
is really relevant is itself a theoretical issue and not one that can be solved by appeal
to plausibility.

Tye also objects that biological mechanisms are actually not preserved in the
elementary-particle simulators, contrary to what I claim. He says,

It is also important to realize (contra Block) that there isn’t really any water in your body
after the infestation. For water could have existed without any elementary-particle people but
“water” could not. So water and “water” differ in their modal properties and thus are not
the same. Likewise, there aren’t really any electrons or protons (but rather groups of tiny people
acting as if they are electrons and protons—people who can change their minds at any time
and thereby cease to function in elementary-particle-like ways). But if there is no water in your
body, if there are no hydrogen ions, no oxygen, no carbon, no protons, no electrons, then the
electrochemical mechanisms are not the same after the elementary-particle-people infestation
as they were before.

But the issue is whether what is important to the electrochemical mechanisms are
preserved. I mention when I describe the elementary-particle simulators that one law
of elementary particles is this: a quark with a certain color together with an antiquark
with the anticolor of that color form a color-neutral meson. And a law of this form
applies to the simulated matter. “Quarks” are related to “color” and “anticolor” and to
“mesons” just as quarks are related to color and anticolor and to mesons. Real neural
firing involves an electrical difference across the cell membrane created by the action
of the sodium-potassium pump. When the neuron fires, sodium channels open and
sodium ions pass through the cell wall, reversing the charge difference between the
inside and the outside of the cell. The corresponding process with quotation marks
placed appropriately is the description of the new weird matter. Quotes or no quotes,
the lawful operation of the biological mechanisms is the same. If it is the operations
of these biological mechanisms that underlie consciousness (and that is my hypoth-
thesis), then consciousness is preserved in the elementary-particle simulators, and not for
historical reasons but for biological reasons.
Note


References


