

# Absent Qualia and the Mind-Body Problem

---

*Michael Tye*

University of Texas at Austin

At the very heart of the mind-body problem is the question of the nature of consciousness. It is consciousness, and in particular *phenomenal* consciousness, that makes the mind-body relation so deeply perplexing. Many philosophers hold that no definition of phenomenal consciousness is possible: any such putative definition would automatically use the concept of phenomenal consciousness and thus render the definition circular. The usual view is that the concept of phenomenal consciousness is one that must be explained by means of specific examples and associated comments.

The explanation typically proceeds along something like the following lines: there is something it is *like* to taste Green Chartreuse, to hear a chainsaw, to smell a skunk, to see the clear, blue sky. Each of these states has a distinctive subjective character or raw “feel” to it. These raw “feels”—qualia, as they are often called—resemble and differ from one another to varying degrees. The subjective “feel” of the experience of red, for example, is more like the subjective “feel” of the experience of orange than it is like the subjective “feel” of the experience of green. Subjective “feels” or qualia are what make the states possessing them phenomenally conscious.

Further illumination is sometimes offered by noting that it is phenomenal consciousness that gives rise to talk of an explanatory gap and

I am grateful to a referee for the *Philosophical Review* for very helpful comments. I am also indebted to Mark Sainsbury both for discussion and for written suggestions.

*Philosophical Review*, Vol. 115, No. 2, 2006

DOI 10.1215/00318108-2005-013

© 2006 by Cornell University

to two famous philosophical hypotheses: the inverted spectrum hypothesis and the absent qualia hypothesis. However, realism about phenomenal consciousness does not require acceptance of either of these hypotheses.

The absent qualia hypothesis is the main focus of the present essay. This hypothesis, as I shall understand it, is the hypothesis that *could* be the case that a system that functionally duplicates the mental states of a normal human being has no phenomenal consciousness (no qualia). I shall say that qualia are absent just in case they are missing under the given condition of functional duplication. This condition is one that, according to functionalists, guarantees mental duplication, including duplication in qualia.<sup>1</sup>

The absent qualia hypothesis can be understood in three ways: a) that absent qualia are nomically possible; b) that absent qualia are metaphysically possible; c) that absent qualia are conceptually or logically possible. Current orthodoxy in the philosophy of mind has it that absent qualia are at least conceptually possible.<sup>2</sup> This is the view of all dualists about phenomenal consciousness and many materialists. I have come to think that orthodoxy is wrong. Proper and full a priori reflection upon the putative case of absent qualia demonstrates that they are impossible. This conclusion has important consequences not only for

1. Here is a brief statement of functionalism, as it is normally understood. Let *T* be a psychological theory (commonsense or scientific) made up of causal generalizations that connect psychological states to one another, to physical inputs, and to behavioral outputs under normal conditions. Form the long conjunction, *C*, of these generalizations and suppose, to simplify things, that in the generalizations, the mental-state terms are all names rather than predicates. Next replace each mental-state name by a different variable and bind each variable by an existential quantifier to form a sentence of the following form:

$$(\exists x_1)(\exists x_2) \dots T(x_1, x_2, \dots).$$

This is the Ramsey sentence for *T*. We can now give the conditions under which an individual is in pain (has pain), for example, according to the functionalist: Where ' $x_k$ ' is the variable replacing 'pain',

$$\text{A subject } S \text{ is in pain if and only if } (\exists x_1)(\exists x_2) \dots [T(x_1, x_2, \dots) \& S \text{ has } x_k].$$

Since there are no mental-state terms left on the right-hand side of this definition (the mental-state names having been replaced by variables), the definition provides a non-circular functional definition of pain as a state that plays a certain causal role in a network of inner states that is tacked down at its periphery by physical inputs and outputs. Likewise for other mental states. For more, see Lewis 1972.

2. In earlier work, I took the view that absent qualia, though conceptually possible, are metaphysically impossible in maximal functional duplicates of normal human beings.

how philosophers should understand consciousness but also for current scientific work on the mind-body problem.

The essay is divided into eight sections. I begin with an example due to Ned Block of a possible system that has seemed to many philosophers to offer clear support for the absent qualia hypothesis. This example is useful not only because of its intuitive appeal but also because consideration of it allows for a clear illustration of just what is relevant to the absent qualia hypothesis and what is not. In section 2, I turn to a debate about absent qualia that took place some twenty-five years ago between Block and Sydney Shoemaker about absent qualia. This debate, in my view, laid the foundation for current orthodoxy on absent qualia. Shoemaker's argument against the absent qualia hypothesis is presented as is Block's critique of it. In section 3, I lay out Shoemaker's subsequent response to Block and in section 4 I find it lacking. Section 5 considers one further inconclusive attempt (in the general spirit of Shoemaker's response) to argue against the absent qualia hypothesis. The next section develops my own argument that, however things may seem initially with respect to systems such as the one presented in section 1, absent qualia really are impossible, indeed conceptually impossible. Shoemaker may have lost the battle, but, in my view, he was on the winning side in the war. Section 7 draws out the implications of my argument for philosophy. Section 8 does likewise for science.

### **1. The China-Body System**

In his well-known essay "Troubles with Functionalism," Ned Block (1980a) offers the following example: imagine that a billion Chinese people are each given a two-way radio with which to communicate with one another and with an artificial (brainless) body.<sup>3</sup> The movements of the body are controlled by the radio signals, and the signals themselves are made in accordance with instructions the Chinese people receive from a vast display in the sky that is visible to all of them. The instructions are such that the participating Chinese people function like individual neurons, and the radio links like synapses, so that together the Chinese people duplicate the causal organization of a normal human brain down to a very fine-grained level. Does this system undergo experiences and feelings? Block himself holds that intuitively the system does not.

3. Block also offers a variant on this example in which the Chinese people are replaced by homunculi. For more on homunculus-headed systems, see section 2 and later.

Functionalists have replied to Block that this assessment is too hasty. Being so much smaller than the China-Body system, we fail to see the forest for the trees and we leap to an unlicensed conclusion about the absence of qualia—one no more to be trusted than the conclusion a tiny extraterrestrial might reach if he happened to materialize inside a human brain and concluded that the pulses of electricity running down the huge number of pathways surrounding him do not support experience.

If the issue that interests us is whether absent qualia are conceptually possible, this reply is not to the point (and neither is Block's appeal to intuition on behalf of the example). For whether or not this system, if it is actualized, has any feelings and experiences, it seems conceivable that it does not. If this is correct, then it is conceptually possible for a system to duplicate a normal human functionally with respect to mentality and yet lack qualia.

Here is a parallel. Suppose a certain rock *R* is dropped. Will it fall to earth? Of course, it will. The prevailing laws of nature are such that if *R* is dropped, it falls. Still, there is no inconsistency in supposing that the laws themselves are false, that other laws are operative. So, even though it is nomologically necessary that if *R* is dropped, then it falls, still it is conceptually possible that it does not do so.

Correspondingly, in the case of the China-Body system, even if we decide that it is nomically impossible for the system to lack experiences, we have not thereby decided that it is conceptually impossible. After all, it seems that we can imagine the system itself being without any feelings and experiences, just as we can imagine the rock, *R*, floating in the air or moving away from the earth. But if we really can imagine these things, then they are certainly conceptually possible. And that entails that the absent qualia hypothesis, on its weakest reading, is true. In what follows, my interest is in the absent qualia hypothesis on this reading. I shall argue that the hypothesis, so understood, is false.

## **2. Shoemaker's Argument That Absent Qualia Are Impossible and Block's Critique**

Shoemaker's argument (1975), as reconstructed by Block (1980b), is essentially as follows. Suppose

- (1) Absent qualia are possible.

Then

*Absent Qualia and the Mind-Body Problem*

- (2) There could be a state functionally exactly like pain but lacking its phenomenal character (ersatz pain).

Since ersatz pain would have the same causal consequences as real pain, it follows that

- (3) The phenomenal character of pain makes no difference to its causal consequences.

But, by the causal theory of knowledge,

- (4) We have knowledge of the phenomenal character of pain only if it causes our introspective beliefs about pain.

So,

- (5) We do not have knowledge of the phenomenal character of pain.

Patently this is false. So,

- (6) Absent qualia are impossible.

To this argument, Block replies that (3) is false. It could be that the phenomenal character of genuine pain is essential to its having the causal consequences it does even though the same consequences could have been produced in another way by ersatz pain. Consider, for example, a computer that works by electricity. There could be a second computer computationally identical to the first that works hydraulically. The former computer lacks fluid—it is an absent fluid device—but it would be silly to conclude that the presence of fluid in the hydraulic computer makes no difference to its operations. The absent fluid computer is a computational isomorph of the electrical computer—it produces the same outputs as the hydraulic computer via the same computations—but it does so in virtue of states that have electrical properties rather than hydraulic ones.

Block describes a homunculus-headed system to illustrate this point further. Suppose that there is a body externally like your body but internally very different. In this body, the sensory neurons are connected to a bank of lights in a hollow cavity in the head, and the motor neurons become active as a result of buttons being pushed on a console. Within the cavity, there are a number of different groups of little men. Each man has been trained to identify a particular light combination representing an input when the letter indicating his group is displayed on a bulletin board. For his allocated input, the relevant little

man then pushes a particular output button in accordance with a rule he has memorized and changes the letter on the bulletin board, again in accordance with the memorized rule. Even though each man is of very low intelligence and has an exceedingly simple task, the homunculus-headed system as a whole is functionally equivalent to you at the psychological level. That is to say, it is subject to a system of inner states that interact with one another, sensory inputs, and behavioral outputs in just the same ways as your psychological states. But, according to Block, this system need not feel pain when it is in a state functionally like the one you are in when you experience pain. In Block's view, it is reasonable to suppose that the system here lacks qualia. But the possibility of such a homunculus head with absent qualia clearly does not show that our genuine pain does not have *its* causal consequences in virtue of its qualitative feel or character.

### 3. Shoemaker's Reply to Block

Shoemaker (1981) begins his reply by claiming that functionalism can be taken in a weak sense or a strong sense.<sup>4</sup> This part of Shoemaker's discussion can be bypassed for present purposes.<sup>5</sup> The key claim Shoemaker

4. Functionalism, in the weak sense, is the doctrine that each mental state is definable in terms of its causal relations to inputs, outputs, and other mental states, whether or not these other mental states are themselves functionally definable. Functionalism, in the strong sense, is the doctrine that each mental state is definable in terms of its causal relations to inputs, outputs, and other mental states in such a way that no mental vocabulary occurs in the definiens.

5. I should add that, in his reply to Block, Shoemaker distinguishes between two versions of the absent qualia hypothesis, corresponding to the two versions of functionalism distinguished in note 4, and he claims that his original argument was directed only against the weak absent qualia hypothesis (the hypothesis that there could be a system that is a weak functional duplicate of a normal human being but which undergoes no experiences). This then creates a further difficulty for Block's critique, according to Shoemaker, since the homunculus head is supposed to lack mental states generally and thus is not a weak functional duplicate of a normal human. Two comments: (1) Block does say in his essay that the head seems as lacking in thought as qualia; but thoughts typically come clothed in auditory qualia (via associated auditory, linguistic images), and this may be why a doubt about qualia extends to a doubt about thoughts. Moreover, as Block has acknowledged in recent conversation, it is the attribution of qualia to the head that is seriously problematic. In his (present) view, there is no obvious difficulty with the attribution of beliefs, for example. (2) In his reply, Shoemaker revises the argument so as to apply to the strong absent qualia hypothesis. Since the absent qualia hypothesis is normally understood in the strong sense, as is functionalism, in this section I ignore the weak version and assume that the strong absent qualia hypothesis is at issue.

makes with respect to Block's presentation of his original argument is that premise (3) is a distortion. If a state is functionally equivalent to pain at the psychological level, then it must have the same causal consequences with respect to the relevant internal states, inputs, and outputs (those used in the functional definition of pain). But it need not have the same total causal consequences. So, premise (3) needs qualification. The qualification Shoemaker offers is this:

- (3') The phenomenal character of pain makes no difference with respect to those of its causal consequences that would make it possible to distinguish cases of genuine pain from cases of ersatz pain.

Once the premise is stated in this way, it is obvious, according to Shoemaker, that an insuperable epistemological problem arises. For given (3'), in Shoemaker's (1981, 588) words, "my grounds for thinking that my own pains are real and not ersatz can be no better than the grounds that an imitation man would have for thinking the same about his ersatz pains."<sup>6</sup> The revised argument Shoemaker now proposes is simply this:

- (7) The absent qualia hypothesis entails that phenomenal states are unknowable by their subjects.
- (8) We have knowledge of our own phenomenal states.

So,

- (9) The absent qualia hypothesis is false.

Shoemaker buttresses this argument with a second, related argument that appeals to imitation people as follows:

- (10) If the absent qualia hypothesis is true, then there is a possible world *W* in which there are imitation people as well as genuine people.
- (11) In *W*, both sorts of people have beliefs about phenomenal states and thus concepts of such states; and both use words like 'pain' that refer to such states.

6. Here an imitation man is one who is functionally *exactly* like a normal human being but without qualia.

- (12) In *W*, ‘pain’, as used by the imitation people, means the same as ‘pain’, as used by the genuine people; likewise, for the word ‘qualia’ if the imitation people are sophisticated enough to have such a word.

So,

- (13) The pains of the (supposedly) imitation people in *W* are genuine, not ersatz.

So,

- (14) There is no possible world *W* in which there are imitation people as well as genuine people,

and thus,

- (15) The absent qualia hypothesis is false.

In support of (12), Shoemaker says that the situation is like that in which we encounter Martians who are apparently psychologically isomorphic with us, who speak a language that sounds just like English, but whose biochemistry and neurophysiology are very different. If we are entitled to say that the Martian “pains” lack qualia, they are entitled to use the same-sounding words to say the same about us. Patently, they are not so entitled. Neither then are we. Shoemaker (*ibid.*, 575) continues:

I think it is obvious that if we mingled with the Martians, did business with them, struck up friendships with them, talked philosophy with them, and so on, we would use “pain” and the rest as if they applied univocally to them and us, and we would have no inclination to give these words a parochial interpretation. A proponent of the parochial view (the view that in our language words like “pain” and qualia have their reference fixed by states we have and the Martians do not<sup>7</sup>) would have to say that in doing so we would be changing the meanings and references of these words. I suppose that it is not possible to prove that this view is wrong; but it seems to me utterly implausible.

Likewise, according to Shoemaker, for the imitation people we encounter in world *W*.

7. Material in parentheses added from elsewhere in Shoemaker’s essay.



#### 4. Deficiencies in Shoemaker's Response

First, some general remarks about concept possession and belief. It is clear that one could have a concept of red and not have experienced red, as, for example, with the case of a blind person or Frank Jackson's Mary, locked in her black and white room. Likewise, one might have a concept of pain and never have experienced pain. Still intuitively, there is *a* concept of red and *a* concept of pain such that possessing them requires (respectively) knowing what it is like to experience red and knowing what it is like to experience pain. Without these concepts, one could not have the sort of understanding of red and pain that normal humans have via their experiences of red and of pain. Let us call such concepts phenomenal concepts. These concepts are ones exercised in what we may call "phenomenal beliefs." I shall return to the topic of phenomenal beliefs shortly.

Now it is widely held that beliefs *generally* pose no special problem for functionalism. To see why, *prima facie*, beliefs are unproblematic, consider another being who is a functional duplicate of me (not only narrowly but broadly too). Suppose we ask my duplicate where the philosophy department at the University of Texas is. He'll reply, "Waggener Hall," just as I would. If we ask the duplicate whether he is a philosopher, he'll say that he is. If we ask him whether he is going to the colloquium in Waggener due to start in five minutes, he'll nod and head off in the direction of Waggener Hall. Why does he walk off in the direction he does? Surely because he believes that *that* is the direction of Waggener and he wants to hear the colloquium there. His behavior in this instance and indefinitely many others (both linguistic and nonlinguistic) sanctions the attribution of beliefs and desires. For how else are we to explain my duplicate's behavior?

The point is not that it is inconceivable for a being without a mind to behave in an apparently intelligent fashion. Indeed, a being preprogrammed with a huge look-up table of responses to stimuli clearly might behave just as I do without having any beliefs and desires. The point is that in the case of the functional duplicate, there are inner states that function in just the ways beliefs and desires do with respect to one another and behavior, and it seems that our concept of a belief or a desire is the concept of a state that plays an appropriate functional role.<sup>8</sup>

8. This is not to say that beliefs and desires always bring about the effects that are partly constitutive of the relevant roles. A brain placed in a vat and supplied with the appropriate inputs may continue to have beliefs that do not cause any bodily behavior.

So, it seems that my functional duplicate, like me, has beliefs and desires. And these beliefs and desires explain his behavior. The principles of psychological explanation apply to him just as they do to me.

Of course, this is not to deny that there is room for disagreement about just how the attribution of beliefs is to be spelled out functionally. Arguably, the relevant functional roles will bring in teleological factors with respect to what normally causes what and will include scientific generalizations, both wide and narrow, as well as wide commonsense ones; and these cannot all be elucidated a priori. But this does not matter for present purposes. What seems to me clear is that any system that is a *complete* psycho-functional isomorph of me, that is, any system that duplicates my psychological states functionally across the board not only at the manifest, commonsense level but also at the level of science will be subject to beliefs as a matter of conceptual necessity.<sup>9</sup>

What is not so clear, indeed what is highly contentious in the present context, is the claim that such a functional duplicate will be subject to *phenomenal* beliefs. Consider, for example, the phenomenal belief that one is in pain. If it is conceptually necessary that one undergo genuine pain in order to have the phenomenal concept *pain*, then if pain is not functionally definable, as the advocate of the absent qualia hypothesis will insist, neither is the phenomenal belief that one is in pain.<sup>10</sup> Accordingly, for all that Shoemaker has shown, pain and ersatz pain may be held to give rise to different introspective beliefs. Genuine pain causes in me the *phenomenal* belief that pain is present; ersatz pain, it may be urged, causes in my functional duplicate the *nonphenomenal* belief that ersatz pain is present.

On this view, the subject of genuine pain knows that he is in genuine pain since he has the phenomenal belief that he is in pain (a belief

9. Philosophers who accept the absent qualia hypothesis and who hold further that there is something it is like to have a belief (so that beliefs qua beliefs essentially have phenomenal character) will reject this claim. Their view is not at all plausible, however. Where there is phenomenal character associated with a belief, the belief is manifested in a conscious thought and the phenomenology attaches to the thought or to images (typically linguistic) that go along with the thought or in which it is encoded. In the absence of a thought, there is nothing it is like to have a belief. (Even those who do not accept this claim should agree that there is nothing it is like to have a deeply repressed belief. From this alone it follows that beliefs do not essentially have phenomenal character.)

10. Given that the phenomenal belief that one is in pain exercises the phenomenal concept *pain*.

that requires for its truth that its subject is in genuine pain), his belief is true, and it is appropriately warranted (by, for example, a reliable introspective process that takes phenomenal states as input and yields true phenomenal beliefs as output when it is operating properly). Likewise, the subject of ersatz pain knows that he is in ersatz pain. It follows that phenomenal states are not unknowable to their subjects even if the absent qualia hypothesis is true. Shoemaker, then, is not entitled to assume premise (7) in his first revised argument. Further reasons need to be given in support of (7)—reasons that Shoemaker has not supplied.

Turning now to the case of the Martians used in the second argument, we can agree with Shoemaker that we would no doubt use the word ‘pain’ to apply to them. But it is surely conceivable (for all that Shoemaker has shown) that if we are using this word to express a phenomenal concept, we are mistaken in so applying it. Conceivably, it may be insisted, the Martians don’t feel pain. To be sure, they apply the word ‘pain’ to themselves. But there is no logical guarantee that the word, as they use it, expresses a phenomenal concept.

Admittedly, Shoemaker might well reply that his aim was only to show that absent qualia are metaphysically impossible, not that they are conceptually or logically impossible. But why should we accept that it is even metaphysically impossible for the Martians to be using the word ‘pain’ to pick out a nonphenomenal state? Relatedly, why should we concede that the concept the imitation people express by ‘pain’ is the same as the concept the genuine people express by ‘pain’? The imitation people, like the Martians, apply the word ‘pain’ directly to their own inner states from time to time, and they then function as if they are feeling pain. But still, according to the advocate of the absent qualia hypothesis, their concept is one that applies to ersatz pains. So, when the imitation people say things like “I am in pain,” they are *not* expressing beliefs about their own phenomenal states. Shoemaker may find this implausible; but, given that the first-person report, “I am in pain,” as issued by the imitation people, is true, they cannot be charged with making any gross error of a sort that the genuine people are not. It seems to me, then, that premise (11) in Shoemaker’s second argument is open to doubt as is premise (12).

I want next to consider another line of reasoning, similar in spirit to Shoemaker’s imitation people argument. This line of reasoning in the end is unsuccessful, in my view. That is to say, it does not succeed in proving that absent qualia are *impossible*. But reflection upon it may

begin to weaken the grip the absent qualia hypothesis has on many philosophers.

### 5. A Further Look at Phenomenal and Nonphenomenal Belief

Consider again the specific phenomenal concept *pain* and suppose that I have a functional duplicate who lacks any phenomenal consciousness (hereafter *NN*). Since *NN* does not experience pain, the concept he expresses by 'pain' is not the one I express by 'pain'. Precisely how similar to my phenomenal concept is his?

To begin with, *NN*'s concept, like my phenomenal concept *pain*, is plausibly viewed as atomic. This needs a little explanation. One way to explain the notion of an atomic concept is via the language-of-thought view. On this view of concepts, a concept is a mentalese symbol that a thinker is disposed to use in certain ways in thoughts. The concept is atomic if and only if the relevant mentalese symbol is syntactically atomic. This account of an atomic concept is neutral on the further question as to whether the symbol is *definable* in terms of other mentalese symbols. Another more general way to explain the notion of an atomic concept is to say that it is a mental representation of a sort exercised in thought and belief that has no other mental representation as a component part. This account does not preclude the possibility of an a priori analysis of an atomic concept in terms of other concepts. For example, one might hold that the concept *fortnight* is an atomic concept while also holding that it is analyzable via the nonatomic concept *fourteen days*. The difference in these concepts is what explains the informativeness of the assertion that a fortnight is fourteen days (unlike the assertion that a fortnight is fortnight).

With these preliminary points in place, it should be clear that if we suppose that my concept is atomic and *NN*'s is not, we risk introducing psycho-functional differences between the two of us since it could well be that this difference will engender differences in our dispositions to say certain things and draw certain inferences. Let us grant, then, that both concepts are atomic.

Is *NN*'s atomic concept empty? Given that his concept is not a phenomenal concept, it is very hard to see how it could be. For surely the only plausible correctness conditions for his (nonphenomenal) concept are that it applies to a state *S* if and only if *S* is a state that functions just like pain but lacks its phenomenal character (if and only if, that is, *S* is ersatz pain). Since *NN* undergoes ersatz pains, *NN*'s concept is not

empty. When *NN* sincerely reports “I am in pain,” he says something true just as I do when I make the same report.<sup>11</sup>

What further characteristics does *NN*'s concept have? My phenomenal concept *pain* is both atomic and rigid: it is a simple concept that picks out the same phenomenal state in all possible worlds. Furthermore my concept is a *direct* recognitional concept at least to the following extent: when I apply the concept I do not consciously base my application on inference or on evidence that is different from its referent, as I do when I apply the concept *gold*, say.<sup>12</sup>

Correspondingly, *NN*'s concept is an atomic, rigid, and direct recognitional concept (in the above sense). It differs from my concept only in its having a nonphenomenal referent. His concept, however, cannot be a direct recognitional concept of the sort applied by a super-blindsighter (Block 1997) who comes to believe his guesses about observational properties of things presented to him even though he cannot see those things. The latter subject has no idea how he comes to believe the things he does. If asked, he will say that the belief comes to him out of the blue. *NN* is not like *that* any more than I am. In my case, if you ask me on what I base my belief that I am in pain, I will reply “the way my present experience feels.” In saying this, I am not implying that my belief is supported by reasons as such. But I do not believe my belief is entirely groundless. Rather, I believe that its ground is the subjective character of my experience (the very thing to which the phenomenal concept *pain* refers). So, I believe that I apply my concept on the basis of how my state feels.

So, we can now conclude the following about *NN*'s concept: it is an atomic, nonempty, rigid, direct recognitional concept applied, so he maintains, on the basis of how his experience feels.

What holds for this one concept holds for all the other functional counterparts to my phenomenal concepts that *NN* possesses, including the most general phenomenal concept *experience*. Furthermore, just as my phenomenal concepts have a character that leads to the explanatory gap questions, so *NN*'s surrogate phenomenal concepts lead to corresponding questions. That is, just as I can intelligibly ask myself such questions as, “Why should these brain states generate the feeling of pain? Why should they generate any feeling at all?”, so too, by the assumption of

11. For further arguments in support of this claim, see Balog 1999.

12. This characterization of a direct recognitional concept is weaker than that offered in Loar 1997.

functional duplication, *NN* can ask questions formulated in exactly the same way.<sup>13</sup> And *NN* will be puzzled by these questions just as I am.

With all of the above constraints in place on *NN*'s concepts, the idea that they might *still* pick out nonphenomenal entities surely loses much of its intuitive appeal. Is it now really so obvious that the concept *NN* expresses by "pain" *could* be any different from my concept? If not, then since my concept is phenomenal and requires for its possession that I undergo or have undergone the experience of pain, so too does *NN*'s concept make the same requirements on him. *NN*, thus, cannot lack qualia, contrary to the initial assumption.

Still the argument so far does not demonstrate that absent qualia are *conceptually* impossible. For the defender of the absent qualia hypothesis can insist that there is no *proof* here that *NN* has qualia. Nonetheless, it should at least now be clear that the requirement of functional identity contained in the absent qualia hypothesis imposes standards that are not easy to meet and that go well beyond the coarse-grained functional similarities that initially spring to mind.

By way of further illustration of this point, consider the following very early remarks about the possibility of absent qualia by Keith Campbell (1970, 100–101)—remarks that predate the Block-Shoemaker debate:

suppose a being very like us except that instead of feeling a pain when he burns his finger or breaks his toe, he has no locatable sensations at all. He just spontaneously gains a new belief, it just "pops into his head" that he has burned his finger or broken his toe, as the case may be. Call this being an *imitation man*. His awareness of his own body would be like our awareness that the car we are driving in is getting a flat tire. Some change in our body, of which we are not conscious, has as a result that it just pops into our heads that the tire is going flat. Awareness of the kind we have, that our finger is burned, ceases at the end of successful, soothing operations. The bare belief of the imitation man that his finger has been burned could just disappear in the same way, as our belief that the tire is flat evaporates when we change the wheel.

A few pages later, Campbell extends the imitation man example so that all qualia are absent, not just those pertaining to bodily sensation. Even with this extension, however, Campbell's imitation man would not count as one of Shoemaker's imitation people. For, as described above, the imitation man does *not* functionally duplicate the awareness I have that my

13. For more on the explanatory gap, see section 7.

finger is burned. I believe (and thus I will say if asked) that my belief that I have a burning pain in my finger is based on how my finger feels whereas Campbell's imitation man will say no such thing. His "pain" belief, we are told, just pops into his head. So, to allow the possibility of functional similarities as described in the above passage without the presence of qualia does *not* suffice to allow the truth of the absent qualia hypothesis.

In the next section, I turn to the construction of a further argument for the conclusion that it is conceptually impossible for a being who is a *complete* psycho-functional isomorph of a normal human to lack phenomenal consciousness.

## **6. The Real Trouble with Absent Qualia**

Debates about personal identity sometimes proceed by reflection upon thought experiments in which the participants in the debates ask themselves whether, if faced with certain alternatives, they would choose this one or that. For example: if my brain were transplanted into your body and your brain were transplanted into mine and you wanted to survive, then if you knew that one of the two bodies had to be destroyed after the transplants and that the choice would be yours, would you choose to have my original body (now with your brain) or your original body (now with my brain) destroyed? Or: if a copy of my brain and body were made by means of a teletransporter (a device that scans each cell in my brain and body and then creates copies of them elsewhere), and you knew that afterward one of the two brains would be destroyed, if you wanted to survive, which brain would you want to be eliminated?

The argument I develop below proceeds in part in a similar way. I begin with the following principle:

- (P) Necessarily, if family *F* of mental states in being *S* has members that are one-to-one functionally isomorphic with the members of family *F'* of mental states in being *S'*, where *S* and *S'* are themselves psycho-functional duplicates, then exchanging the two families preserves psycho-functional duplication.

(P), in my view, is an a priori truth. Some clarifications follow. First, it should be noted that the envisaged exchange is at the level of mental states. So, exchanging physical realizers of the mental states comprising the families is relevant to (P) only if it brings with it an exchange of the mental states themselves. Secondly, the functional isomorphism to

which (P) adverts demands an isomorphism with respect to the entire pattern of causal relations the relevant mental states bear to other mental states, inputs, and outputs. Duplication of inputs and outputs alone for the salient mental states does not suffice. Thirdly, an exchange of the families of mental states,  $F$  and  $F'$ , is to be understood to occur only if  $S$  comes to be subject to the members of  $F'$  where previously he had been, or would have been, subject to the corresponding members of  $F$  (with as few other changes as possible) and conversely for  $S'$ . Finally, (P) is not to be read as implying that after the exchange  $S$  and  $S'$  must function at the mental level just as they did before. That seems obviously false. The claim rather is that necessarily, after the exchange of the two families of mental states,  $S$  and  $S'$  continue to be mental functional duplicates *whether or not* they function as they did before.

Those who oppose functionalism may respond that (P) is all well and good if the families,  $F$  and  $F'$ , have members with functional role definitions (wide or narrow). But what if  $F$  and  $F'$  are families of phenomenal states? In this case, it may be tempting to think that (P) does not apply. This would be a mistake, however. (P) applies even if functionalism is false.

To see this, consider the case of two people whose color experiences are phenomenally inverted. These people (let us grant for present purposes) function psychologically in the same way. Suppose that overnight an operation is performed upon the visual system of each so that afterward each has color experiences with the same phenomenal character as the other before the operation. Upon awakening, neither is a functional duplicate of his earlier self. For each now says such things as "This is very weird: the sky today looks yellow; fire engines look green." Even so, after the exchange of phenomenal color experiences, each is a functional duplicate of the other.

The general point here is that with the exchange of a family of phenomenal states for another inverted family *within* a given individual, if the members of the two families are indeed one-to-one functionally isomorphic, then immediately after the exchange, as far as the functioning of the members of the new family go, that individual is psycho-functionally just as another individual would have been at the corresponding time, had he been subject initially to the inverted family and then had it replaced by the first family. So, if the two individuals are psycho-functional duplicates before the exchange, they must be such duplicates afterward too. The cogency of (P) in such a scenario shows that it does not presuppose functionalism. But (P) does draw an important a



priori link between the mental and the functional that is exploited in the argument that follows.

The point just made need not be restricted to the exchange of whole families of phenomenal states, it is worth noting. If the focus is shifted to the exchange of individual functionally isomorphic mental states,  $M$  and  $M'$ , an individualized version of (P) still holds.

As an illustration, suppose that the operation is a *partially* inverting one that reverses neuronal activity only in the opponent processing channel associated (in normals) with the experiences of red and green. Here the normal member of the pair of people will say, after the operation, that grass looks red (as in the last case) and further that oranges look greenish-yellow. But these are just what the originally inverted member of the pair will say after the operation too. Again, after the exchange, the two people are psycho-functional duplicates.

I turn next to my thought experiment. Imagine that a complex device has been constructed with dual head-caps, each with probes protruding from its inner surface—probes that painlessly penetrate the skull when the head-cap is worn. These head-caps are connected to one another and other supporting machinery and computers in such a fiendishly clever way that when a switch is thrown, tiny robots enter the brains of the two people wearing the head-caps through the probes and make various internal changes with lightning speed before withdrawing back up the probes. The result of these changes is that there is a partial exchange of phenomenal states and nonphenomenal states between the two people, an exchange that continues after the head-caps are removed. This exchange is such that were I and my functional duplicate,  $NN$ , to agree to undergo the exchanger operation (as it comes to be called), I would lose all my phenomenal states, *other than those that are phenomenal memories*, and I would have them replaced by corresponding ersatz phenomenal states.  $NN$  would lose all his ersatz phenomenal states, *except those that are ersatz phenomenal memories*, and he would have them replaced by corresponding phenomenal states.

Clearly, the exchanger operation is conceptually possible if it is conceptually possible for me to have a functional duplicate lacking qualia.<sup>14</sup> In my own case, it suffices for me simply to imagine hearing the

14. It might be objected that the exchanger operation is conceptually impossible on the grounds that it is conceptually impossible for internal changes alone to result in an exchange of phenomenal and nonphenomenal states. However, even those who hold (as I do) that it is metaphysically possible for microphysical duplicates to differ

switch being flipped after I don the head-cap and then being struck very forcefully a little later by my lack of phenomenal consciousness with respect to my body and the world around me, while retaining memories of my past psychological life even though I still find myself functioning in very much the same way as I did before. With this change, of course, I would not be a total functional duplicate of my earlier self since I would notice the difference in my psychological life.

Moreover, I would not be entirely without phenomenal consciousness. For in the envisaged scenario, the memories I would have on the basis of which it is manifest to me that, for example, I no longer have bodily sensations and perceptual experiences via my senses are phenomenal memories. They are memories of how things looked to me and smelled and felt; and it is via these memories that I say such things as “Grass did look green to me; but it doesn’t look any way to me now.” Such memories have phenomenal character: they recreate in imagistic form the earlier experiences, and thus there is something it is like to undergo them.

Now ask yourself the following question: if I were offered the opportunity to undergo the exchanger operation, and I were convinced that the operation was completely safe, and further I knew that the operation would not be reversed at some later date, would I choose to have the operation or not?

In my own case, I would decline the exchanger operation out of hand. Why? It is not that I do not know what it would be like after the operation (to put the matter in Nagelian [1974] terms) and thus that I lack a firm basis for a comparison of my mental life before and after. I do know what it would be like: assuming that the operation is successful, as far as my phenomenal states other than phenomenal memories go, it would be like nothing at all, for *ex hypothesi* afterward I would have *no* phenomenal consciousness with respect to the world around me and my body. Why then? The answer is that in undergoing the operation, I would lose something of *value* to me, something I strongly do not want to lose, something I *care* about very much: to wit, my perceptual and bodily phenomenal consciousness. Clearly, the operation would diminish my mental life. Conversely, if I think about the situation of *NN* and reflect upon whether he would gain something of value were he to undergo an

---

phenomenally should allow that it is *conceptually* possible that the internal changes the tiny robots make have the envisaged results if it is conceptually possible for me to have a functional duplicate lacking qualia in the first place.

operation that replaced his ersatz phenomenal states (other than his ersatz phenomenal memories) with genuine phenomenal ones, I have no doubt that he would. I am certain that the operation would enrich his mental life, not diminish it. He would gain something worth having that he lacked before the operation, something of more value than his ersatz phenomenal life. After all, he would now have “technicolor phenomenology”;<sup>15</sup> and technicolor phenomenology is *better* than no phenomenology at all.

Of course, if I were unfortunate enough to suffer constant, intense pain, I might take a different view. I might well decide to undergo the operation because the negative value of the experience of pain outweighed the positive value of phenomenal consciousness generally. But note that now it seems clear that if this is the view I take as I reflect upon whether *NN* would gain something of value if he were to undergo the reverse operation, I must conclude that he would not. After the operation, *NN* will undergo constant, intense pain experiences. So, if I care about *NN*, surely I would try to persuade him *not* to allow himself to participate in the exchanger operation. His mental life afterward would be worse than before even if he functioned in much the same way. To simplify the discussion, I shall ignore the “awful pain” scenario for myself and my duplicate in what follows.

The next point I want to stress is that if *NN* and I were to undergo a partial exchange of the sort I have described, not only would I have lost something of value, I also would *think* of myself as having lost something of value to me. I would *think* of my mental life as diminished. This would be manifest to me via the comparison provided by my phenomenal memories.

The transformation envisaged here for me via the imagined exchange is structurally like that which obtains in the first stage of an intrapersonal inverted spectrum case in which I undergo an operation that crosses wires in my visual system so that my color qualia are inverted. Upon awakening from the operation, I am no longer functionally the same as I was before. Now, relying upon my memories of my earlier experiences, I am aware that a striking change has occurred. Correspondingly, in the above intrapersonal absent qualia scenario.

In the inversion case, with the passage of enough time, many years later the memories disappear and I now call grass “green,” lemons “yellow,” and so on, just as everyone else does; still later, as the case is

15. The term is McGinn’s (1989).

usually developed, I bang my head and suffer amnesia with respect to the original operation without anything else in my psychological life being affected. At this stage, I am functionally the same as I was before the operation. But there is a definite phenomenal difference. In a counterpart development of the intrapersonal absent qualia case, after the initial stage described above, I imagine that I lose all my memories, including my phenomenal ones of my life before the fateful operation. Now, it seems, I am a full-fledged zombie replica of my earlier self.

It should be emphasized that whether this last stage of the intrapersonal zombie replica case really is conceptually possible—whether, that is, it is conceptually possible for me to become a full-fledged zombie—is something on which presently my argument need take no stand. For the moment, my aim is simply to show that once we suppose that it is conceptually possible for me to have a complete functional duplicate without qualia, then we are committed to accepting that it is conceptually possible for the two of us to undergo a *partial* exchange of phenomenal and nonphenomenal states of the sort described above (whether or not a total exchange is conceptually possible).

Now given (P), if *NN* and I have our phenomenal and nonphenomenal states exchanged, other than those that are phenomenal memories in my case and ersatz phenomenal memories in his, then we must be functional duplicates of one another after the exchange, just as we were functional duplicates of one another before. But surely that will not be the case. As noted above, I will compare my current mental states with those I remember using my phenomenal memory and find my current mental states lacking. *NN*, however, will have a very different cognitive response. Using his ersatz phenomenal memory (memory that functions just like phenomenal memory but with products that lack any phenomenal character), surely he will not take a negative view of his situation. Indeed, surely he will appreciate that he has gained something important, something very special he did not have before.

Here is another way to make the point. Conceive of *NN* initially as lacking qualia. Then conceive of him as gaining them with respect to his bodily sensations and the world as perceived around him, while still having only ersatz phenomenal memories of his past qualia-free life. To conceive of him thus entails conceiving of him as not only gaining something of value after the exchange but also, via the use of memory and introspection, as appreciating that his psychological situation is *better* than it was before. Certainly, it seems incoherent to suppose that on

the basis of a comparison between his past mental life, as provided by memory, and his current mental life, he could suppose that he has lost something of value, that his mental life has been diminished.

Suppose you deny this. Suppose you say that *NN* will indeed have gained something of value he did not have before and I will have lost it, but there will be no difference in what we think or believe after the operation. I will believe that I have lost something important *and so will he*. It is just that his belief, unlike mine, will be false. Then how is it that *NN* manages to form the false belief that he has lost something important? This belief cannot just suddenly pop into *NN*'s head out of thin air; it is formed directly on the basis of introspection and memory, just as my belief is. How is it that *NN* manages to go so badly wrong, given that his mechanisms of introspection and memory are working as well as mine? Furthermore, if you take the view that *NN*'s belief is false while mine is true, then, it seems to me, you are left with an unjustifiable asymmetry in the case. For why not say instead that *my* belief that I have lost something of value is false and *NN*'s belief is true? Lastly, if you hold firm to the claim that *NN*'s belief that he has lost something of value is false—that he has really gained something of value just as I have—then surely we finally do lose *all* grip on the idea that his mental life could be any different from mine at all.

Of course, if *NN* and I are genuine functional duplicates, then there *cannot* be a difference in our cognitive reactions of the sort I have been insisting upon (a difference that will manifest itself in a difference in verbal behavior, for example). But there must be such a difference, I have argued. That is what the above reasoning compels us to conclude. The contradiction reached here shows that it is *not* conceptually possible for me to have a complete functional isomorph who undergoes merely ersatz phenomenal states. Necessarily, any system that functionally duplicates me is phenomenally conscious. The absent qualia hypothesis, therefore, is false even on its weakest interpretation.

What we have here, I suggest, is a compelling argument that closes the gap that the argument of the last section seemed to leave open. There it was noted that the advocate of the absent qualia hypothesis could insist that however strange and far-fetched it might seem, the possibility remains that I have a functional duplicate, lacking in phenomenal consciousness, with concepts that pick out merely ersatz phenomenal states. Now we see what we should have suspected anyway in light of the earlier reflections: that for myself and any psycho-functional duplicate,

there is no genuine difference in the two of us with respect to the presence of phenomenal consciousness. Shoemaker was right: absent qualia are impossible. More strikingly, they are *conceptually* impossible.<sup>16</sup>

If this conclusion is correct, then we can know a priori that phenomenal consciousness has a functionally determined nature. This, it is worth noting, is compatible with the claim made in section 5 that phenomenal concepts are atomic (including the concept *phenomenal consciousness*). For, as noted there, the question of whether a concept is atomic is independent of the question of whether it is definable (or analyzable). The conclusion just reached is also compatible with the claim that phenomenal concepts are direct recognitional concepts. To see this, it suffices to appreciate that what is required of a direct recognitional concept, as understood earlier, is only that it be a concept, the recognitional use of which is not consciously based upon either inference or underdetermining evidence of a quasi-perceptual sort. Supposing that phenomenal concepts meet this characterization is compatible with adopting any of a number of different proposals concerning what goes on *tacitly* as such concepts are applied via introspection, including various functionalist theories.<sup>17</sup>

There remain two worries I want to address. First, it might be objected that if my argument against absent qualia is sound, then the

16. It is worth noting here that an argument similar in spirit to that given above can be mounted for the conclusion that absent qualia are conceptually impossible in functional duplicates without appealing to (P). To see this, note first that if *NN* and *I* are functionally alike, then if *I* find what *I* have (phenomenal consciousness) valuable, then *NN* will find what he has (ersatz phenomenal consciousness) valuable in the same way; for a state's being found valuable is a functional feature of it. Further, *I* do find phenomenal consciousness valuable—that is shown by my reaction to the possible operation described earlier. So, *NN* must find ersatz phenomenal consciousness valuable. But if he does, then this refutes the supposition that *NN* has only ersatz phenomenal consciousness lacking in value.

It might be replied that *NN* could be mistaken in finding what he has valuable. But *I* have already argued against a very similar suggestion (see p. 159). The points made there apply *mutatis mutandis* in the present context. A further point worth making and noted by a referee is that *NN* can't find what he has *as* valuable as *I* find what *I* have even though, given the functional isomorphism, he does find what he has of value. This is because while *NN* is content with what he has, if only he were to experience what *I* have (as he will, if he undergoes the exchanger operation), he would see that what he has is decidedly inferior. Not so the other way around. But this difference is impossible, given that *NN* and *I* are functional duplicates of one another after the exchanger operation. (This line of reasoning brings in (P) again.)

17. For more on functionalism and tacit knowledge, see Braddon-Mitchell and Jackson 1996.

China-Body system described in section 1 could not fail to be phenomenally conscious. So, why does it seem to us that it might lack qualia? If it is not conceptually possible for the China-Body system to fail to undergo experiences, why does it seem to us that we can imagine it without them?

Leaving to one side the point that on some ways of developing the China-Body example, there fails to be *wide* functional duplication (and thus that the system fails to be a complete functional isomorph of a normal human being), the most convincing reply, it seems to me, is simply this. Take the case of a mathematical conjecture that I have not managed to prove and which, as far as I know, is open. To me, it may well seem that the conjecture could be false. I have no proof to the contrary. As far as I am concerned, there is no contradiction in supposing that the conjecture is false. So, it may seem to me that I can imagine the conjecture being false. But suppose that the conjecture actually has a straightforward proof which I have failed to uncover. Then, there is an *a priori* deduction that the conjecture is true, and thus the conjecture is, in the relevant sense, conceptually necessary even though it seems to me that I can imagine it failing. Likewise, if I have not gone through the reasoning presented above. In that case, it may well seem to me that I can imagine a homunculus-headed system that duplicates a normal human functionally and yet lacks qualia. But in reality this is not conceptually possible.

Secondly, a worry may be raised with respect to my dialectical strategy in this section.<sup>18</sup> The argument I have given is essentially a *reductio*. I assume that absent qualia are conceptually possible and then, relying upon the intuition that phenomenal consciousness is something of value—something I care not to lose—a contradiction is derived. But if the conclusion that absent qualia are conceptually impossible is correct, then why trust the intuition? Intuitions one has when something one thought to be conceptually possible turns out to be conceptually impossible are dubious at best.

My initial reply is that the “intuition” in this case is something I know to be true *a priori* simply by reflecting upon my experience and what matters to me. It does not rest upon the assumption that absent qualia are conceptually possible. In rejecting that assumption, my prior intuition about the value of phenomenal consciousness is not threatened; for in the context of my argument, the intuition that is needed is

18. This worry is due to a referee.

really no more than the thought that if I were to lose my phenomenal states other than my phenomenal memories and have them replaced by ersatz phenomenal states, then I would have lost something I value. And that seems to me undeniable. A further point is that the objection purports to cast doubt on the intuition in the case that the conclusion of my argument is correct. But if the conclusion is admitted to be correct by the objector, then the intuition is not needed. After all, the point of giving an argument is to try to persuade those who *disagree* with its conclusion.

If, as I have argued, absent qualia are conceptually impossible, then they are metaphysically and nomically impossible too.<sup>19</sup> I turn next to the implications of this position for philosophical work on the mind-body problem.

### **7. Consequences for the Philosophy of Mind**

If it is conceptually impossible that a system which is a functional duplicate of a normal human being with respect to mentality lacks phenomenal consciousness, then it is knowable a priori that phenomenal consciousness has a functionally determined nature. Consistent with this claim, it could be that the nature is narrow, bringing in no facts about the environment or historical setting. It could be that the nature is wide. It could be that the nature is a commonsense, functional one fixed by the causal generalizations of commonsense psychology. It could be that the nature is scientific, fixed by some pertinent range of the causal generalizations of scientific psychology. It could even be that the nature is scientific and cognitively closed from us; for conceivably the relevant range of causal scientific generalizations employ concepts that we will never be able to acquire, given the structure of our minds. So, a version of the “mysterian” view of consciousness (McGinn 1989) remains open.

Another possibility—and the one I myself accept—is that some version of representationalism is true for phenomenal consciousness. According to representationalists (Dretske 1995; Tye 1995, 2000; Jackson 2003), phenomenal consciousness is a matter of undergoing states whose representational contents meet certain further conditions. So long as these conditions are themselves fixed by functional organization, and so long as the relevant kind of representational content is fixed in like man-

19. This conclusion is one that Nagel himself has recently endorsed, notwithstanding his denial of it in earlier work. See his 1998.



ner, nothing in the argument of this essay is incompatible with representationalism. To be sure, representationalists sometimes write in ways that suggest that their view is strongly opposed to functionalism. But this is only true if functionalism is construed on one particular model or other. Once functionalism is allowed to be either wide or narrow and to bring in historical and/or environmental facts, there need be no conflict with representationalism.

If my argument is correct, then the type-identity theory cannot be correct for phenomenal consciousness, however. And neither can the view that phenomenal consciousness has a nonfunctional, spiritual nature. So, while our philosophical options still remain sufficiently open to allow choice between commonsense functionalism, scientific functionalism (either mysterian or nonmysterian), and representationalism, some important theories are no longer available to us.

There are consequences for other arguments and problems concerning phenomenal consciousness. For example, if absent qualia are conceptually impossible, then there cannot be a world that is physically just like the actual world in all respects and thus that contains creatures who are microphysical duplicates of normal human beings, where these creatures lack phenomenal consciousness. So, arguments for dualism (Chalmers 1996) based upon the conceptual possibility of such a zombie world must fail. Most notably, the problem of the explanatory gap is seen to lose much of its force. This can be brought out as follows.

We all have a grasp of the subjective or phenomenal character of such states as the visual experience of orange, the feeling of anger, the sensation of being about to sneeze. Our grasp of what it is like to undergo these and other experiential states is supplied to us by introspection. We also have a grasp, albeit an incomplete one, of what goes on objectively in the brain and the body. But there is, it seems, a vast chasm between the two. Presented with the current physical story of the objective changes that occur when such-and-such subjective feelings are experienced, we have the strong sense that the former does not fully explain the latter, that the phenomenology has been left out. We naturally ask: What is so special about *those* physical goings-on? Why do they feel like *that*? Indeed, why do they feel any way at all?

Compare this case with that of brittleness or digestion, say. Once one learns that in brittle things, the intermolecular bonding forces are weak (much weaker than in resilient things), one immediately grasps that brittle things break easily. Having been told the physical story, to ask: "Yes, but why are things with low intermolecular bonding forces brit-

tle? Why shouldn't such things not be brittle?" is to show a conceptual confusion. One who responds in this way simply does not understand the ordinary notion of brittleness. What it is for something to be brittle is for it to be disposed to break easily. Once the forces between molecules are sufficiently weak, they can be broken apart with little effort, and thus objects with such internal bonding forces are disposed to break easily, that is, to be brittle.

Similar points apply in the case of digestion. Upon learning that there are enzymes in the alimentary canals of human beings that break down food and convert it into energy, only a failure to grasp that the word 'digestion' means (roughly) *internal process whose function is to convert food into energy* could lead one to ask: "Why does the action of these enzymes in humans generate digestion? Why shouldn't the enzymes turn food into energy in the absence of digestion?"

In the case of phenomenal consciousness, however, the corresponding questions remain, it is usually held, even for those who are not suffering from any conceptual deficiency. One who has a full understanding of the term 'experience' and who is fully apprised of the physical facts, as we now know them, can still coherently ask why so-and-so physical processes feel any way at all or why they feel the way they do. In this case, it seems that as far as our understanding goes, *something important is missing*. Herein lies the famous "explanatory gap" for consciousness.<sup>20</sup>

I have argued that our concept of phenomenal consciousness is a concept whose application in our own case we can know a priori to be determined by human functional organization. Although this does not fully close the explanatory gap—in particular, it does not explain why, once a certain functional organization is in place, an experience is like *that* and not like something else—we can no longer coherently ask why, *given* that organization, phenomenal consciousness is present. To do so would be like asking why, given that a device functions so as to trap mice, it is a mousetrap. Just as it is inconceivable that a mousetrap is not present once a device is at hand that functions appropriately, so too it is inconceivable that phenomenal consciousness in humans does not go along with human functional organization. The difference between the two cases is that we know a priori that the concept *mousetrap* is the concept of a device that functions so as to take free mice as input and deliver

20. The expression "explanatory gap" was coined by Joseph Levine. See his 1983.

trapped mice as output, but we do not know a priori (or so it seems to me) which functional role is the relevant one for phenomenal consciousness. So, we may well wonder which part of our functional organization does the trick. To this extent there is indeed an explanatory gap as far as the presence of phenomenal consciousness goes; and if mysterians are right, there will always be one. But the *form* of the relevant explanation is now clear. The underlying physical states realize certain privileged higher-level functional states, and the latter a priori necessitate phenomenal consciousness.

### **8. Consequences for Scientific Theories of Consciousness**

Recent work in science on consciousness has divided into two camps. In the one camp, there are those who see consciousness, including phenomenal consciousness, as neurobiological in nature. Scientists (for example, Kanwisher [2001], Driver and Vuilleumier [2001]) who take this view typically point to the impressive array of correlations that have been discovered between various neural activations and perceptual experiences of one sort or another, for example, experiences of faces, houses, letters, words. The most general (and best-known) correlation here is the one Crick and Koch (1990) have alleged to obtain between the neuronal oscillation of 40 MHz and phenomenal consciousness. On the basis of this hypothesized correlation, Crick and Koch assert that phenomenal consciousness just is a neuronal oscillation of 40 MHz.

Falling into the second camp are those scientists who take a “brain-neutral” view of consciousness. It is not always clear just which sort of consciousness is the target of the proposed scientific theory, but the most popular “brain-neutral” proposal is that consciousness is a matter of global availability of information, that is, the availability of information to many different consumers. As Dehaene and Naccache (2001, 1) put it:

An information becomes conscious . . . if the neural population that represents it is mobilized by top-down attentional amplification into a brain-scale state of coherent activity that involves many neurons distributed throughout the brain. The long-distance connectivity of these “workspace neurons” can, when they are active for a minimal duration, make the information available to a variety of processes including perceptual categorization, long-term memorization, evaluation and intentional action. We postulate that this global availability of information through the workspace is what we subjectively experience as a conscious state.

Admittedly, in this passage, Dehaene and Naccache write of activity in the *neuronal* workspace, and this is not brain-neutral, on a literal reading. But the presumption of scientists in this camp is that neuronal activity is simply the way in which consciousness is implemented in the human biological system. Consciousness itself is global accessibility. This is also the view taken by Dan Dennett (1991) when he asserts that consciousness is fame in the brain or cerebral celebrity.<sup>21</sup> Again, there is no commitment to an architecture of neurons as necessary to consciousness.

The contrast between the two camps is at bottom the contrast between type physicalism and functionalism.<sup>22</sup> Type physicalists hold that consciousness and its various species, for example, the experience of pain, the feeling of a tickle, the visual experience of red are identical with the physical realizers or implementations of so-and-so functional roles. Functionalists, by contrast, hold that consciousness and its various species are identical with functional-role states or properties.

Much of the recent debate in science has focused on two issues: whether the correlations between various conscious states and neuronal states are really robust<sup>23</sup> and how the functionalist proposal is best developed. My interest here is not in the details of these debates but rather in the more general question of their fruitfulness as they pertain to the issue of the nature of phenomenal consciousness. And here the argument of this essay has something to say. For if my reasoning earlier is correct, then the type-identity theory cannot be correct, and thus scientists would do best to concentrate their efforts on the second issue.

Of course, it may be that some of the scientists who apparently endorse type physicalism for consciousness really only hold that the relevant physical states realize the conscious phenomena in human beings in much the same way as, for example, a certain molecular pattern in glass realizes brittleness. To the extent that this is the case, the two camps I have distinguished are not really at odds, and the argument of this essay has no implications for scientific work. But insofar as scientists are trying to get at the nature of consciousness, and phenomenal consciousness in particular, then my argument entails that “brain-neutral,” broadly functionalist proposals are the only ones on the right track. Silicon implementations of the underlying functional organization of consciousness

21. See also Dehaene, Kerszberg, and Changeux 1998.

22. For a detailed discussion of the two camps, see Block 2001.

23. There is evidence that the neural correlate of visual experience in normal vision also occurs without visual consciousness in masking and priming.

must be conscious even though they do not share our biological nature. What matters about matter, I maintain, at least as far as consciousness is concerned, is only how it functions.

## References

- Balog, K. 1999. "Conceivability, Possibility, and the Mind-Body Problem." *Philosophical Review* 108: 497–528.
- Block, N. 1980a. "Troubles with Functionalism." In *Readings in the Philosophy of Psychology*, 1:268–305. Cambridge, MA: MIT Press.
- . 1980b. "Are Absent Qualia Impossible?" *Philosophical Review* 89: 257–74.
- . 1997. "On a Confusion about a Function of Consciousness." In *The Nature of Consciousness*, ed. N. Block, O. Flanagan, and G. Güzeldere, 375–415. Cambridge, MA: MIT Press.
- . 2001. "Paradox and Cross Purposes in Recent Work on Consciousness." *Cognition* 79: 197–219.
- Braddon-Mitchell, D., and F. Jackson. 1996. *Philosophy of Mind and Cognition*. Oxford: Blackwell.
- Campbell, K. 1970. *Body and Mind*. Garden City, NY: Anchor Books.
- Chalmers, D. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Crick, F., and C. Koch. 1990. "Towards a Neurobiological Theory of Consciousness." *Seminars in the Neurosciences* 2: 263–75.
- Dehaene, S., M. Kerszberg, and J. P. Changeux. 1998. "A Neuronal Model of a Global Workspace in Effortful Cognitive Tasks." *Proceedings of the National Academy of Sciences USA* 95: 14529–34.
- Dehaene, S., and L. Naccache. 2001. "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework." *Cognition* 79: 1–37.
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.
- Dretske, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press, Bradford Books.
- Driver, J., and P. Vuilleumier. 2001. "Perceptual Awareness and Its Loss in Unilateral Neglect and Extinction." *Cognition* 79: 39–88.
- Jackson, F. 2003. "Mind and Illusion." In *Minds and Persons*, ed. Anthony O'Hear, 251–71. Cambridge: Cambridge University Press.
- Kanwisher, N. 2001. "Neural Events and Perceptual Awareness." *Cognition* 79: 89–113.
- Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly* 64: 354–61.
- Lewis, D. 1972. "Psychophysical and Theoretical Identification." *Australasian Journal of Philosophy* 50: 249–58.

- Loar, B. 1997. "Phenomenal States." In *The Nature of Consciousness*, ed. N. Block, O. Flanagan, and G. Güzeldere, 597–616. Cambridge, MA: MIT Press.
- McGinn, C. 1989. "Can We Solve the Mind-Body Problem?" *Mind* 98: 349–66.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83: 435–50.
- . 1998. "Conceiving the Impossible and the Mind-Body Problem." *Philosophy* 73: 337–52.
- Shoemaker, S. 1975. "Functionalism and Qualia." *Philosophical Studies* 27: 291–315.
- . 1981. "Absent Qualia Are Impossible—A Reply to Block." *Philosophical Review* 90: 581–99.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press, Bradford Books.
- . 2000. *Consciousness, Color, and Content*. Cambridge, MA: MIT Press, Bradford Books.