

# The Problem of Qualia

By BRENT SILBY  
Department of Philosophy  
University of Canterbury

Copyright (c) Brent Silby 1998

[www.def-logic.com/articles](http://www.def-logic.com/articles)

One of the major concerns for the contemporary philosophy of mind involves the problem of qualia. The word qualia refers to our subjective experience of the world and includes the properties of our experience that cannot be located in the world external to our minds. For example, the ineffable *feel* of a blue experience when one looks at the sky, or the pain one *feels* when one is stuck with a pin. These sensations are the essence of our experience and yet cannot be pointed to in the external world.

This paper is based upon a talk I gave to the 1998 cognitive science class and will follow the same structure of that talk. To start, I will describe the problem of qualia and will show that it poses a real problem for physicalist and functionalist theories of the mind. Further highlighting the problem, I will talk about androids. These entities of the future are functionally identical to humans, however it is claimed that if they existed, they would have no qualia or any conscious experience whatsoever. The final section of this paper will be concerned with Frank Jackson's Knowledge argument. The Knowledge argument is a very powerful thought experiment, which is supposed to show that physicalism about the mind is false.

## The Problem of Qualia

The existence of qualia is a problem for cognitive science and philosophy of mind because it seems that our subjective sensations of experience only exist within the mind. According to the traditional view, qualia are characterised as being:

- (1) ineffable
- (2) intrinsic
- (3) private
- (4) directly apprehensible in consciousness.<sup>1</sup>

Sensations of colours, pains and sounds seem very real to us and yet they cannot be found in the world external to our minds. If we wish to ask what science can tell us about the colour *blue*, all we learn is that there is a certain wavelength of electromagnetic radiation that is being reflected off

various objects. There is no blue, just radiation. The question we are left with is: where is the *blueness* if it is not in the outside world? One possible answer would be to describe the visual process like this:

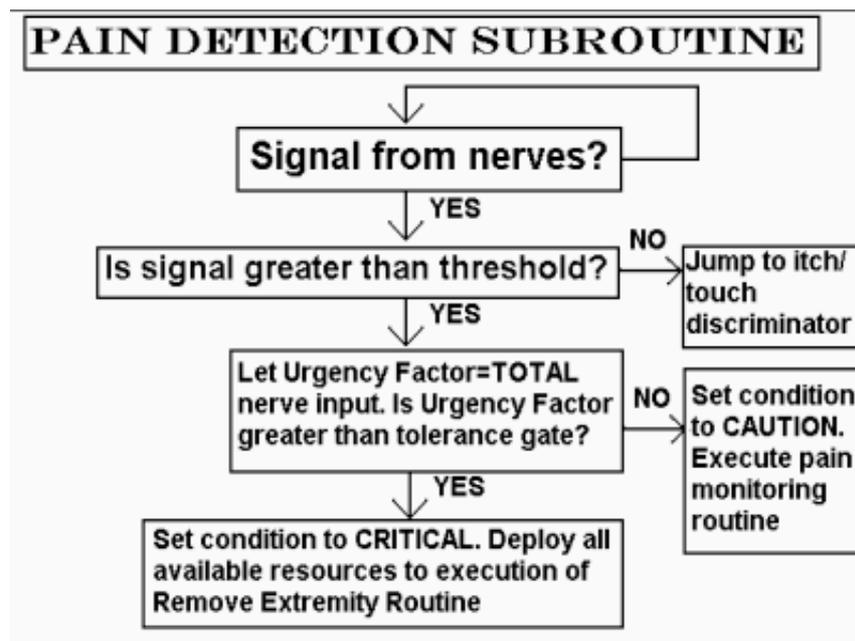
*When light enters our eyes, various frequencies of radiation cause the retina's cones to activate. This initiates an electrochemical cascade by which signals are sent to the visual cortex and a neural state is set up. A connected 'box' is then filled with a representation of the colour, which can then be examined by other parts of the brain in order to produce behavioural output.*

Now, it certainly *feels* like this is going on in our brains but there are problems with this picture. Daniel Dennett suggests that it would be a waste of resources for the brain to work in this fashion. Once a sensory discrimination has been made, there is no need for it to be made again by other parts of the brain. In other words, the visual cortex has discriminated the colour so the colour representation box is unnecessary. The brain does not need to contain a box within which a colour is represented in order to be examined by other parts of the brain. Nature has given the brain a visual system that can make the colour discriminations by itself. But this brings us back to the question: Where is the *blue* that we experience? There is no blue in the outside world, and there is no box filled with blue colour inside our heads so why do we have the blue sensation? It is this question that has led many philosophers to the opinion that qualia does not supervene on the functional organisations of the brain. This is to say that even if we had a complete understanding of the brain's functional organisation, we would not be able to account for qualia because they cannot be captured by a functionalist account of the mind. So, we could imagine a situation in which qualia do not exist and yet the brain's functional states remain the same. Many thought experiments have been devised to show this point. They usually involve imagining an entity that is functionally (and sometimes physically) identical to us, but has no conscious experience whatsoever.

### **What would it be like to be an Android?**

Computer technology is progressing at an astonishing rate and is likely to continue to grow. Humans have designed computer programs that can distinguish colours with great accuracy but we do not suppose that these computer programs experience any colour qualia. But what about computers of the future? A great deal of effort is being put into researching artificial intelligence and we can imagine that at some point in the future, humans will successfully design intelligent androids.<sup>2</sup> Now, if we imagine that these androids are behaviourally identical to humans we can ask the question: do they experience qualia? Would the designers of these androids need to build qualia into their programming? Presumably not. If we consider pain, we can imagine that the android's software would

contain a pain handling routine similar to this:



The pain handling routine this flow chart depicts is simplified but the point is clear. Once instantiated in an android, it will monitor sensory input and will jump to other sub-routines depending on the strength of the input. There are no qualia included in this system and it would appear that including *feeling* would not be necessary. The system is effective as it is. An android's brain could be full of similar routines that handle all of its sensory input and it would survive in the world perfectly all right but without the *feel* of experience. Of course the android has been designed to imitate human behaviour, so if we asked it a question about its experience, it would respond in much the same way as we would. The point is that we could not tell from its outward behaviour that there is nothing it is like for it to be an android. Now, what could we say about qualia if we could show that our brains work in a similar fashion to the android brain? Qualia would seem to be an unnecessary bonus that escapes our functionalist picture of the mind. Of course we could question the idea that our brains work in that way but the point should be clear. The android scenario shows us that beings without qualia could be behaviourally indistinguishable from beings with qualia.

Let's alter the thought experiment by imagining that neuroscientists have a complete understanding of how the brain works. We can imagine that they supply cybernetics researchers with a complete wiring diagram of the human brain. The cyberneticists could then construct an android whose brain is functionally identical to the human brain. Its silicon brain would work in exactly the same way as the human brain and its behaviour would be identical to human behaviour. Such an entity is known as a 'Functional Isomorph'. The only difference between its brain and a human brain is what materials it happens to be constructed of. Now, here's the important question: Would the android be conscious and experience qualia? Intuitively, many people would say no. They might claim that the android

brain is not the right type of system to give rise to conscious experience. On the other hand, many people (myself included) may claim that if the android's brain was functionally identical to our human brains, it would be conscious (in so far as we are conscious). But that might imply that *any* system would be conscious if it was functionally identical to the brain. Ned Block points out that it would be possible to get every person in China to simulate a neuron and allow them to simulate neural connections by equipping each of them with cell phone links. We can imagine that if this was carried out effectively, China would be functionally the same as the human brain. But, says Block, this system would not be conscious.<sup>3</sup> Intuitively it is tempting to agree with Block. It certainly seems that the population of China (like an android brain) is the wrong type of system to exhibit a collective consciousness. But surely we could make the same claim about the human brain. We could say that intuitively it seems implausible that this spongy organic brain would be the right type of thing to give rise to conscious experience. But it does.<sup>4</sup> This does not prove that an android brain or the population of China would be conscious, but it is an effective counter to the claim that these systems would not.

The android and Chinese nation examples of 'absent qualia' are supposed to convince us that the functional organisation of the brain cannot account for qualia. These thought experiments appeal to intuition and are easily answered: If the functional organisation of the brain can give rise to qualia, then why not the same structure instantiated in a silicon system? The absent qualia supporters have two possible replies to this question:

- (1) Qualia and states of consciousness are special properties that cannot be explained by the physical sciences. (This leads to a Dualist approach to consciousness.)
- (2) Qualia require that the functional system upon which they supervene, be constructed from a certain type of physical material. (Biological neurons as in the brain.)

Frank Jackson's famous knowledge argument was put forward in the early 1980's and suggests that reply 1 may be the option to take. The knowledge argument suggests that qualia do not supervene on the physical structure of the brain and thus, cannot be explained in physical terms. Jackson's argument carries a strong intuitive force and has many supporters. I will now outline the argument and address some of the standard criticisms of it.

### **The Knowledge Argument**

Jackson asks us to imagine a woman named Mary who is brought up (from birth) in a black and white room. Throughout her life in this colourless environment, Mary reads many black and white books and learns all the laws of physics. As time goes by, Mary becomes an expert in

neurophysiology and of the functional roles that brain states play in the process of colour vision. Mary's knowledge of the physical and functional organisation of the brain becomes complete to the point that there is nothing that she does not know. But, says Jackson, even with her complete physical knowledge of the brain, Mary does not know *everything* there is to know about the brain because she does not know what it is like to see a colour. Jackson believes that when Mary leaves her room for the first time and experiences her first colour, she would learn something new about the world. For this reason, Jackson claims that the physicalist picture of the brain does not capture everything there is to know about the mind. The qualia are left out. The natural conclusion to draw is that the physicalist story of the brain (and more importantly, the mind) is false. Mary knew *every physical* fact about the world, yet she did not know *everything* about the world.

Jackson's knowledge argument is certainly intuitive. It simply seems to be the case that upon leaving her room, Mary would learn something new about the way the world is. Furthermore, this knowledge could not be predicted by her complete physical knowledge of the brain.

### **Dennett's Reply:**

Daniel Dennett asks us to imagine Mary leaving her black and white room and looking at the sky for the first time. "Ah yes", she says, "that's exactly what I thought blue would look like." Dennett suggests that this is exactly what would happen.<sup>5</sup> According to Jackson's story, Mary knows *everything* about the physical world and this includes everything there is to know about the neurophysiology of the brain. Dennett believes that if Mary knew *absolutely everything physical* about the world and the brain, she would know exactly what to expect when she had her first colour experience. It follows that Mary would not be surprised when she saw the colour red for the first time. She would have predicted all the neural events that would occur upon encountering the colour red and she would be able to make an inference as to what sort of experience she would have. It does not matter that Mary has not learned about colours in the usual way because "Mary is not your usual person."<sup>6</sup>

Dennett's reply to the knowledge argument has attracted a lot of support. Dennett is trying to convince us that if it were possible to know every physical fact about the world, then we would know what colours looked like before we happened to experience them. This line of thought is counter-intuitive and has not convinced everyone, but this is to be expected because it is impossible to imagine what a complete knowledge of physics would be like. It would be difficult for us to say what Mary would or would not be able to anticipate upon leaving her room. Robert Van Gulick suggests that it is at least possible that Mary would know what

to expect when she leaves her room.<sup>7</sup> There is, therefore, a way of answering the knowledge argument before discussion needs to take place.

### **Know-That / Know-How Reply:**

Lawrence Nemirow suggested that the knowledge argument draws on a distinction between *propositional* knowledge (knowing *that*) and *procedural* knowledge (knowing *how*). According to this reply, Mary gains no new knowledge of facts or propositions about the world. What she gains upon leaving her room are new skills and practical abilities.<sup>8</sup> This is to say she learns *how* to recognise the phenomenal properties that her complete physical knowledge of the world had predicted the existence of. The difference between these two types of knowledge can be understood by considering one's knowledge of riding a bicycle. We can construct a complete list of propositions that tell us the rules involved in riding a bicycle, but reading this list will not tell us everything about *how* to ride a bicycle because procedural knowledge is an ability. Now Mary, before leaving her room, knows all the facts and propositions about the world and has a complete understanding of what is involved in colour vision but she lacks some procedural knowledge. She does not know *how* to recognise colours.

This reply to the knowledge argument suggests that upon leaving her room, Mary gains no new facts or propositions about the world. I do not agree with this claim. It may be true that Mary gains new abilities, but surely that is not all she gains when she looks at the coloured world for the first time. It simply seems to be the case that when she leaves her room, Mary learns something new about what the world is like, and this involves propositional knowledge.<sup>9</sup> Now, because Mary's complete knowledge of the physics involved in colour vision falls under the label of propositional knowledge, I do not believe that making a distinction between these two types of knowledge offers a refutation of Jackson's knowledge argument.

### **Churchland's Reply:**

One of the most interesting replies to the knowledge argument was offered by Paul Churchland in his paper "Knowing Qualia: A Reply to Jackson".<sup>10</sup> In this reply, Churchland shows that the Knowledge Argument is invalid. For clarity, I will state the Knowledge Argument here:

- (1) Mary (before her release) knows *everything physical* about other people.
- (2) Mary (before her release) does not know *everything* about other people.

Therefore

(C) There are truths about other people that escape the physicalist story.

Churchland describes a possible system by which the human visual system comes to recognise colours.<sup>11</sup>

In the human visual cortex, there is a region known as V4. When an infant human first sees a colour, a neural pattern is set up within V4 and is reactivated when the infant is exposed to that colour again. Similar patterns are set up when the infant is exposed to other wavelengths of light. In abstract terminology, the V4 centre has partitioned itself into several labels. Each one of these labels is a neural pattern which represents a different colour. Now, during the life of the individual, whenever a colour is experienced, the synaptic weight of the sensory input reactivates the colour's label in V4. Thus the visual cortex has identified and 'labelled' the wavelength of light that is present in the visual field. The important point to note here is that these colour recognition labels must be set up at a very early age and require exposure to colour. Effectively, people have to learn how to see colours. Of course in Mary's case these labels will only be suitable for representations of black and white, and shades of grey because that is all she was exposed to when she was young. Unfortunately she is now too old for new patterns to form in the V4 centre of her visual cortex, so when she leaves her room, she will be unable to see any colours. Her brain will process the environmental input in terms of what it already knows - black and white.

In this story Mary is missing much more than an ability, she is missing a neural processing structure. So, Jackson's premise (2) seems to be true. There is something that Mary does not, and indeed cannot know about other people. This story gives an entirely physical account of visual processing, and as such Mary knows all about this process. In fact, given her complete knowledge, Mary should also be aware of her own representational defect as described above. Thus, it seems that Jackson's premise (1) is true.

Paul Churchland offered the above story as a logical possibility for the process of colour vision. Whether or not it is true remains an empirical question. The important point for Churchland is that it shows a possible situation in which Jackson's conclusion does not follow from his premises and as such, Jackson's argument is invalid.

## **Conclusion**

The possibility of absent qualia in the case of android brains and the Chinese nation thought experiments suggest that it is possible for systems to be functionally identical to the human brain and yet have no qualia. I think that the problem with such arguments lies in the fact that they offer

no empirical support for the claims they make. They rely mainly on intuition and can therefore be answered by people who have different intuitions. It is not at all clear to me that androids or other functional isomorphs would not have qualia as we do.

Frank Jackson's knowledge argument also relies largely on intuition but I think it does so in a more appealing way. It is easier for us to imagine Mary's situation than it is for us to imagine the mind of an android and so the argument carries a stronger intuitive force. According to Dennett, the problem with the Knowledge argument is that if we could truly imagine what Mary's knowledge of the world would be like, we would realise that she *would* know what to expect when seeing colours for the first time. This would support physicalism and provide an answer to the mystery of qualia. To further refute the knowledge argument, Churchland has given us reason to believe that the argument is invalid. This was done by describing a physically possible way in which the visual system works. This would mean that because of her upbringing, Mary would be neurologically incapable of experiencing colours. It is for this reason that I believe that the knowledge argument fails to refute physicalism about the mind and qualia.

## References

Block. N, "Inverted Earth" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997.

Braddon-Mitchell. D and Jackson. F, *Philosophy of Mind and Cognition*, Blackwell Publishers, 1996.

Chalmers. D, *The Conscious Mind*, Oxford University Press, 1996.

Churchland. Paul, "Knowing Qualia: A Reply to Jackson" in *The Nature of Consciousness*, edited by Ned Block, Owen

Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997

Dennett. D, *Consciousness Explained*, Penguin Books, 1993.

Dennett. D, "Quining Qualia" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan and Guven Guzeldere, Massachusetts Institute of Technology, 1997.

Van Gulick. R, "Understanding the Phenomenal Mind: Are We All Just Armadillos? Part 1: Phenomenal Knowledge and Explanatory Gaps" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997.

1 Daniel Dennett, "Quining Qualia" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan and Guven Guzeldere, Massachusetts Institute of Technology, 1997, Page 622.

2 Literature on this subject often refers to robots but I prefer to use the term android. This is because androids, unlike robots, are defined as being similar to humans in appearance and abilities.

3 Ned Block, "Inverted Earth" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997.

4 David Chalmers, *The Conscious Mind*, Oxford University Press, 1996, Page 251.

5 Paul Churchland also offers this reply against the knowledge argument, but I will look at Dennett's version, which is contained in - Daniel Dennett, *Consciousness Explained*, Penguin Books, 1993, page 399.

6 Daniel Dennett, *Consciousness Explained*, Penguin Books, 1993, page 401.

7 Robert Van Gulick, "Understanding the Phenomenal Mind: Are We All Just Armadillos? Part 1: Phenomenal Knowledge and Explanatory Gaps" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997, page 560.

8 Robert Van Gulick, "Understanding the Phenomenal Mind: Are We All Just Armadillos? Part 1: Phenomenal Knowledge and Explanatory Gaps" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997, page 560.

9 David Braddon-Mitchell and Frank Jackson, *Philosophy of Mind and Cognition*, Blackwell Publishers, 1996, page 131.

10 This paper appears in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997, page 571.

11 This is a somewhat shortened version of the story. The full version is contained in Churchland's paper "Knowing Qualia: A Reply to Jackson" in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere, Massachusetts Institute of Technology, 1997, page 572.