

# Q

**qualia.** What it is like to have an experience. Qualia include the ways things look, sound, and smell, the way it feels to have a pain, and, more generally, what it is like to have experiential mental states. ('Qualia' is the plural of 'quale'.) Qualia are experiential properties of sensations, feelings, perceptions, and, more controversially, thoughts and desires as well. But, so defined, who could deny that qualia exist?

Although the existence of subjective experience is not (or anyway should not be) controversial, 'quale'—which is more clearly a technical term than 'subjective experience'—is more often used by those who are inclined to reject the common-sense conception of subjective experience. Here is a first approximation to a statement of what is controversial: whether the phenomenology of experience can be exhaustively analysed in intentional, functional, or purely cognitive terms. Opponents of qualia think that the phenomenology of an experience can be exhaustively analysed in terms of its representational or intentional content ('representationism'); or that the phenomenology of experience can be exhaustively analysed in terms of its causal role ('functionalism'), or that having a subjective experiential state can be exhaustively analysed in terms of having a state that is cognitively monitored in a certain way or accompanied by a thought to the effect that I have that state. If we include in the definition of 'qualia' the idea that the phenomenology of experience outruns such intentional, functional, and cognitive analyses, then it is controversial whether there are qualia.

This definition of 'qualia' is controversial in a respect that is familiar in philosophy. A technical term is often a locus of disagreement, and the warring parties will often disagree about what the important parameters of disagreement are. Dennett (1988), for example, assumes that it is of the essence of qualia to be intrinsic (in the sense of atomic, unanalysable, and non-relational), private (in the sense that any objective test would miss the target), incorrigible (to believe one has one is to have one), and non-physical. Dennett says there are no qualia. Hence the title of his paper, 'Quining qualia'. ('To Quine', according to the tongue-in-cheek Philosophical Lexicon, [www.blackwellpublishing.com/lexicon/](http://www.blackwellpublishing.com/lexicon/), is 'to deny resolutely the existence or importance of something real or significant'.) Of course, Dennett is free to use 'qualia' as

he likes, but a defender of a scientific approach to qualia (the point of view of the author of this entry) will prefer a definition of 'qualia' that allows that science can investigate qualia, that qualia may turn out to be physical, and even that we may discover aspects of introspective beliefs about one's qualia can be mistaken. Indeed, I do not see that a scientific approach can rule out in advance that we could discover, empirically, that qualia are intentional, functional, or cognitive. So I prefer to define 'quale' as an aspect of subjective experience that cannot be shown by a priori or other armchair means to be intentional, functional, or cognitive. This is an epistemic rather than a metaphysical conception of a quale (see Block 2002). Importantly, there is nothing in the conception of qualia that I am advocating that is incompatible with the claim that a quale is a physical state, just as heat is molecular kinetic energy and light is electromagnetic radiation. An empirical reductionist thesis about qualia is legitimate—what is not legitimate is an armchair reductive analysis of qualia (e.g. in functional, representational, or cognitive terms).

1. Absent qualia
2. The inverted spectrum
3. Inverted Earth
4. The knowledge argument
5. The function of qualia

## 1. Absent qualia

Qualia are often discussed with respect to certain thought experiments that purport to demonstrate the falsity of representationism, functionalism, and cognitivist approaches.

One of these thought experiments involves 'absent qualia'. If human beings can be described computationally, as is assumed by the research programme of cognitive science, a robot could in principle be built that was computationally identical to a human. But would there be anything it was like to be that robot? Would it have qualia? (See Block 2002, Shoemaker 1975, 1981, and White 1986.) If it had no qualia, the claim that the nature of qualia is to be found in their computational or functional role is mistaken, and functionalism is false. Further, if a zombie robot could have states that represent the properties of the world that our experiences represent (shape, size, colour, etc.) then representationism is false. (Representationism

## qualia

is the view that the qualitative character of an experience is identical to its representational content.) Some thought experiments (Block 1980) have appealed to oddball realizations of our functional organization, e.g. the economy of a country. If an economy can share our functional organization, then our functional organization cannot be sufficient for qualia. Many critics simply bite the bullet at this point, saying that the oddball realizations do have qualia. Lycan (1996) suggests thinking of these roles as requiring evolution and as involving the details of human physiology. Economies do not have states with the right sort of evolutionary 'purpose', and their states are not physiological. But the evolution stipulation is incompatible with the very plausible doctrine that qualia supervene on the brain: that is, no difference in qualia without a neural difference. If we are ever able to manufacture a brain like ours, it will also lack evolutionary 'purpose', but to the extent that it is physically just like our brains, it will serve as the neural basis of similar experience. On the idea of including physiological roles in our functional characterization of mental states, note that this will make the definitions so specific to humans that they will not apply to other creatures that have mental states. (Block (1980) calls this the chauvinism problem.) Further, this idea violates the spirit of the functionalist proposal, which, being based on the computer analogy, abstracts from hardware realization.

### 2. The inverted spectrum

Imagine that we could 'invert' colour processing in the visual system in one of two twins at birth. (See Palmer (1999) for a discussion of a variety of kinds of 'inversion', some of which are more likely to be possible than others.) If the twins grow up using colour terms normally, we might hypothesize that things they both call 'red' look to one the way things they both call 'green' look to other, even though they are functionally (and therefore behaviourally) identical. This inverted spectrum hypothesis would be further confirmed if the brain state that one twin has on seeing red things is the same as the brain state that the other has on seeing green things (and the same for other complementary colours)—assuming supervenience of qualia on the physiological properties of the brain. The relevance of this version of the inverted spectrum hypothesis to qualia is that a physicalist view of qualia can be used to challenge functionalism (qualia are causal roles) and representationism (qualia are representational contents).

For when one twin is looking at a red thing and the other is looking at a green thing, they can have two states that are physically and qualitatively the same, but are functionally and representationally different, one representing red and the other green. And if both twins are looking at the same object, they may have states that are functionally

and representationally the same but qualitatively different. (I assume that all parties can agree on which things are green, which things are red, etc.)

There is a natural reply. Notice that it is not possible that the brain state that I get when I see things we both call 'red' is *exactly* the same as the brain state that you get when you see things we both call 'green'. At least, the *total* brain states cannot be the same, since mine causes me to say 'It's red', and to classify what I am seeing as the same colour as blood and fire hydrants, whereas yours causes you to say 'It's green', and to classify what you are seeing with grass and Granny Smith apples. Suppose that the brain state that I get when I see red and that you get when you see green is X-oscillations in brain area V4, whereas what I get when I see green and you get when you see red are Y-oscillations in area V4. The functionalist says that phenomenal properties should not be linked to brain states quite so 'localized' as X-oscillations or Y-oscillations, but rather to more holistic brain states that include tendencies to classify objects together as the same colour. Thus the functionalist will want to say that my holistic brain state that includes X-oscillations and your holistic brain state that includes Y-oscillations are just *alternative realizations* of the same experiential state. So the fact that red things give me X-oscillations but they give you Y-oscillations does not show that our experiences are inverted. The defender of the claim that inverted spectra are possible can point out that when something looks red to me, I get X-oscillations, whereas when something looks green to me, I get Y-oscillations, and so the difference in the phenomenal aspect of experience corresponds to a local brain state difference. But the functionalist can parry by pointing out that this difference has only been demonstrated intrapersonally, keeping the larger brain state that specifies the roles of X-oscillations in classifying things constant. He can insist on typing brain states for *interpersonal* comparisons holistically. And most friends of the inverted spectrum are in a poor position to insist on typing experiential states locally rather than holistically, given that they normally emphasize the 'explanatory gap' (Levine 1993), the fact that no one has the faintest idea how to explain why the neural basis of an experience is the neural basis of that experience rather than some other experience or no experience at all. So the friend of the inverted spectrum is in no position to insist on local physiological individuation of qualia. At this stage, the defender of the inverted spectrum is stymied.

One move the defender of the possibility of the inverted spectrum can make is to move to an intrapersonal inverted spectrum example. Think of this as a four-stage process. (1) You have normal colour vision. (2) You have colour signal inverting devices inserted in your retinas or in the lateral geniculate nucleus, the first way-station behind the retina, and red things look the way green things

used to look, blue things look the way yellow things used to look, etc. (3) You have adapted, so that you naturally and spontaneously call red things 'red', etc., but when reminded, you recall the days long ago when ripe tomatoes looked to you, colourwise, the way Granny Smith apples do now. (4) You get amnesia about the days before the lenses were inserted. Stage 1 is functionally equivalent to stage 4 in the relevant respects, but they are arguably qualia inverted. And we can add that the subject starts off with X-oscillations when he sees green but the shift induced by the inverters makes green things produce Y-oscillations. So we have an inverted spectrum over time. The advantages of this thought experiment are two. First, the argument profits from the force of the subject's testimony at stages 2 and 3 for qualia inversion. The critic of the intersubjective inverted spectrum says X-oscillations and Y-oscillations are alternative realizations of the same quale, but that is a less plausible move here, since the subject himself says at stage 3 that the state he has now when he sees grass (Y-oscillations) is quite different from the state he used to have on seeing grass (X-oscillations), and there is no reason to suppose this changes at stage 4. Second, the four-stage set-up forces the opponents to say which stage is the one where the inversionist argument goes wrong (see Shoemaker 1981, Block 1990). Rey (1993) attacks stage (3) and White (1995) attacks stage (4).

Nida-Rümelin (1996) notes that, in one kind of genetic colour-blindness, the pigment of the long-wave sensitive cones (L-cones) replaces the normal pigment of the medium-wave sensitive cones (M-cones). Another kind of colour-blindness is the converse: the normal M pigment replaces the L pigment. If a person were to have both kinds of genetic abnormality, they would have the M pigment in the L cones and the L pigment in the M cones. (No such person has been found, but we can predict that there are 14 such persons for every 10,000 males.) Nida-Rümelin argues that such a person might be a natural case of spectrum inversion.

Why, an opponent might ask, is the inverted qualia argument against functionalism any more powerful than a corresponding inverted qualia argument against physicalism? After all, it might be said, one can imagine particle-for-particle duplicates who have spectra that are inverted with respect to one another (Chalmers 1996). But though particle-for-particle duplicates with inverted spectra are in some sense imaginable, they are ruled out by the highly plausible principle, mentioned above, that qualia supervene on physical constitution (i.e. no difference in qualia without a physical difference). By contrast, the form of the inverted spectrum that I have been describing does not flout supervenience: two people can be functionally identical while yet having different physical realizers of those functional states (your red-induced state is realized the same way my green-induced state is).

### 3. Inverted Earth

In discussing the inverted spectrum, I emphasized its relevance to functionalism. There is an interesting variant that is more relevant to representationism. The variant depends on a thought experiment, Inverted Earth (Harman 1982, Block 1990). Inverted Earth is a planet that differs from earth in two relevant ways. First, everything is the complementary colour of the corresponding earth thing. The sky is yellow, the grasslike stuff is red, etc. (To avoid impossibility, we could imagine, instead, two people raised in rooms in which everything in one room is the complementary colour of the corresponding item in the other room.) Second, people on Inverted Earth speak an inverted language. They use 'red' to mean green, 'blue' to mean yellow, etc. If you order paint from Inverted Earth, and you want yellow paint, you fax an order for 'blue paint'. The effect of both inversions is that if you are drugged and kidnapped in the middle of the night, and colour inverters are inserted behind your eyes (and your body pigments are changed), you will notice no difference if you are switched with your counterpart on Inverted Earth.

Now consider the comparison between you and your counterpart on Inverted Earth. We could run the story with the counterpart being your identical twin who was fitted with an inverter at birth and raised on Inverted Earth, or we could think of the counterpart as you after you have been fitted with the inverter and switched with your twin and have been living there for a long while. Imagine you and your counterpart gazing at your respective skies, saying 'How blue!'. Your brains can be stipulated to be molecular duplicates, *not just locally speaking but holistically speaking*. You are looking at different colours, but because of the inverter you have the same overall brain state. So very plausibly, your experiences are exactly the same. And this is plausible independently, since, as the story goes, you notice no difference when you are fitted with the inverter and placed in a niche on Inverted Earth. But the representational content of those experiences is different: your experience represents blue, the colour of the earth sky, whereas your counterpart's experience represents yellow, the colour of the Inverted Earth sky, the colour that they all call 'blue' (meaning yellow). Same qualia but different representational content. And the same set-up argues for the converse case as well. If you are looking at the blue sky on earth while your counterpart is looking at an Inverted Earth lemon (which, you will recall, is actually blue), you will have experiences that represent the same colour, blue, but are phenomenally different. Same representational content, different qualia. And both cases are challenges for representationism.

There are two major advantages of this case over the inverted spectrum thought experiment: (1) the twins' brains are *total* replicas (when they gaze at their respective

## qualia

skies), so there is no potential conflict of holism vs. localism, (2) in the version in which your twin is just you after emigration, there is no need for the adaptation that was required in the intrasubjective inverted spectrum thought experiment.

However, there is a compensating difficulty that derives from the fact that there is a kind of adaptation in the environment rather than in the head. Let me explain. Consider the intrasubjective case. The strength of the intrasubjective inverted spectrum case lies in the personal testimony in stage 3 that grass looks the way blood used to look colourwise. The weakness lies in the fact that the subject has undergone a strange adaptation process that perhaps impugns his judgements. In the case of the subject being moved from earth to a niche on Inverted Earth, when the subject says things look the same now as they always did, he is believable because he undergoes no phenomenal adaptation process. However, we can reasonably wonder whether his experience on Inverted Earth really does represent the sky as yellow. We can agree that the natives all around him represent the sky as it really is, that is, as yellow, but our subject after all grew up (and evolved) using *that very experience* to represent blue. This is one of the objections made in Tye (2000).

One response is that the representational content of experience is not fixed by evolution or early experience. If we find out that your grandparents were 'swamp-people' who came together from particles in the swamp and did not evolve at all, we would not thereby have shown that your experience does not have the normal representational and qualitative content. Also, if I come to develop a Bush recognitional capacity but then move to a place where Bush's twin brother is the president, I don't misrecognize twin Bush forever. Eventually, my recognitional capacity readjusts. My visual experience represents Bush as Bush. So why should we suppose that the colour represented by the subject's early experience on earth makes it forever misrepresent the colour of the sky on Inverted Earth? (I have only been able to touch on issues that are discussed in much more detail in Lycan 1996 and Tye 2000.)

### 4. The knowledge argument

The inverted spectrum, Inverted Earth, and absent qualia arguments are critiques of functionalist and representationalist views of qualia. The final thought experiment to be discussed here is directed as much at physicalism as at functionalism. Jackson's (1986) Mary is raised in a black-and-white environment in which she learns all the functional and physical facts about colour vision. None the less, when she ventures outside for the first time, she learns a new fact: what it is like to see red. So, the argument goes, what it is like to see red cannot be a functional or physical fact. Dennett (1991) objects that perhaps she

could have 'figured out' which things are red; but that is beside the point for two reasons. The question is whether she knows what it is like to see red, not which things are red. And does she know it simply in virtue of knowing all the functional and physical facts about colour vision?

A natural objection is that physicalism is not concerned with what Mary knows, but rather what qualia are, what their metaphysical nature is. Someone could know all the physical facts about molecular kinetic energy changes but not know about heat changes—if the person does not know that heat is molecular kinetic energy. But the analogy does not apply, since we can suppose that Mary *does* know that what it is like to see red is something physical and even exactly what state of the brain it is. That will not keep her from learning a new fact when she learns what it is like to see red.

Lewis (1990) denies that Mary acquires any new knowledge-*that*, insisting that she only acquires knowledge-*how*, abilities to imagine and recognize. But others have emphasized that what Mary learns seems very propositional: she could even express it as: 'What I learned is that *this* is what it is like to see red.' Further, as Loar (1990) points out, the knowledge she acquires can be expressed in embedded contexts. For example, she may reason that, if this is what it is like to see red, then what it is like to see red is more enjoyable than what it is like to be slapped in the belly with a wet fish.

Here is a different (and in my view more successful) objection to Jackson (Horgan 1984, Peacocke 1989, Loar 1990, Papineau 1993, van Gulick 1993, Sturgeon 1994, Perry 2001): what Mary acquires when she sees red for the first time is a new phenomenal *concept*. This new phenomenal concept is a constituent of genuinely new knowledge—knowledge of what it is like to see red. But the new phenomenal concept picks out *old* properties, properties picked out by physical or functional concepts that she already had in the black-and-white room. So the new knowledge is just a new way of knowing old facts. Before leaving the room, she knew what it is like to see red in a third-person way; after leaving the room, she acquires a new way of knowing the same fact. If so, what she acquires does not rule out any possible worlds that were not already ruled out by the facts that she already knew, and the thought experiment poses no danger to physicalism.

### 5. The function of qualia

It is often supposed that qualitative character has all sorts of wonderful functions, promoting flexibility, creativity, and even making recursive models. Block (1995) suggests that the main function of qualitative character is promotive access to mechanisms of short-term memory, perceptual categorization, reasoning, and decision making. Qualia are red flags that representations wave at intelligent processors. The intelligent processing that results is a

product of three things: the preconscious mechanisms that determine what representations are to acquire qualitative character, the post-conscious mechanisms that actually do the intelligent processing, and finally, the qualitative character that helps to make the representations accessible to the intelligent mechanisms. To give the qualitative character *itself* the credit for creativity, flexibility, etc. is like giving the printing press the credit for the ideas that are printed. NB

- Block, N. (1980). 'Troubles with functionalism'. Reprinted in Block N. (ed.), *Readings in the Philosophy of Psychology*, Vol i.
- (1990). 'Inverted earth'. In Tomberlin, J. (ed.), *Philosophical Perspectives*, 4.
- (1995). 'On a Confusion about the Function of Consciousness'. *Behavioral and Brain Sciences*, 18.
- (2002). 'The harder problem of consciousness'. *Journal of Philosophy*, 99/8.
- Byrne, A. (2001). 'Intentionalism defended'. *Philosophical Review*, 110.
- Chalmers, D. (1996). *The Conscious Mind*.
- Dennett, D. (1988). 'Quining qualia'. In Marcel A., and Bisiach, E. (eds.), *Consciousness in Contemporary Society*.
- (1991). *Consciousness Explained*.
- Davies, M., and Humphreys, C. (1993). *Consciousness*.
- Harman, G. (1982). 'Conceptual role semantics'. *Notre Dame Journal of Formal Logic*, 23/2.
- Horgan, T. (1984). 'Jackson on physical information and qualia'. *Philosophical Quarterly*, 34.
- Jackson, F. (1986). 'What Mary didn't know'. *Journal of Philosophy*, 83.
- (1993). 'Armchair metaphysics'. In O'Leary-Hawthorne, J., and Michael, M. (eds.), *Philosophy in Mind*.
- Levine, J. (1993). 'On leaving out what it is like.' In Davies and Humphreys (1993)
- Lewis, D. (1990). 'What experience teaches'. In Lycan, W. (ed.), *Mind and Cognition*.
- Loar, B. (1990). 'Phenomenal properties'. In Tomberlin J. (ed.), *Philosophical Perspectives: Action Theory and Philosophy of Mind*.
- Lycan, W. (1996). *Consciousness and Experience*.
- McGinn, C. (1991). *The Problem of Consciousness*.
- Nida-Rümelin, M. (1996). 'Pseudonormal vision: an actual case of qualia inversion?' *Philosophical Studies*, 82.
- Palmer, S. (1999). 'Color, consciousness, and the isomorphism constraint'. *Behavioral and Brain Sciences*, 22/6.
- Papineau, D. (1993). 'Physicalism, consciousness and the antipathetic fallacy'. *Australian Journal of Philosophy*, 71.
- (2002). *Thinking about Consciousness*.
- Peacocke, C. (1989). 'No resting place: a critical notice of *The View from Nowhere*'. *Philosophical Review*, 98.
- Perry, J. (2001). *Knowledge, Possibility and Consciousness*.
- Rey, G. (1993). 'Sensational sentences switched'. *Philosophical Studies*, 70/1.
- Shoemaker, S. (1975). 'Functionalism and qualia'. *Philosophical Studies*, 27.
- (1981). 'Absent qualia are impossible: a reply to Block'. *Philosophical Review*, 90/4.
- Sturgeon, S. (1994). 'The epistemic view of subjectivity'. *Journal of Philosophy*, 91/5.

Tye, M. (2000). *Consciousness, Color and Content*.

Van Gulick, R. (1993). 'Understanding the phenomenal mind: are we all just armadillos?' In Davies and Humphreys (1993)

White, S. L. (1986). 'Curse of the qualia'. *Synthese*, 68.

— (1995). 'Color and the narrow contents of experience'. *Philosophical Topics*, 23.