In S. Guttenplan (ed) *A Companion to Philosophy of Mind*, Blackwell: Oxford, 1994

QUALIA

Qualia include the ways things look, sound and smell, the way it feels to have a pain; more generally, what it's like to have mental states. Qualia are experiential properties of sensations, feelings, perceptions and, in my view, thoughts and desires as well. But, so defined, who could deny that qualia exist? Yet, the existence of qualia is controversial. Here is what is controversial: whether qualia, so defined, can be characterized in intentional, functional or purely cognitive terms. Opponents of qualia think that the content of experience is intentional content (like the content of thought), or that experiences are functionally definable, or that to have a qualitative state is to have a state that is monitored in a certain way or accompanied by a thought to the effect that I have that state. If we include the idea that experiential properties are not intentional or functional or purely cognitive in the definition of `qualia', then it is controversial whether there are qualia.

This definition of `qualia' is controversial in a respect 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, for example, has supposed in some of his writings that it is of the essence of qualia to be non-relational, incorrigible (to believe one has one is to have one) and to have no scientific nature (see Flanagan, 1992, p 61). This is what you get when you let an opponent of qualia define the term. A proponent of qualia ought to allow that categorizations of them (beliefs about them) can be mistaken, and that science can investigate qualia. I think that we ought to allow that qualia might be physiological states, and that their scientific nature might even turn out to be relational. Friends of qualia differ on whether or not they are physical. In my view, the most powerful arguments in favor of qualia actually presuppose a physicalistic doctrine, the supervenience of qualia on the brain. (See PHYSICALISM).

Perhaps the most puzzling thing about qualia is how they relate to the physical world. Sometimes this is put in terms of the explanatory gap, the idea that nothing we know or can conceive of knowing about the brain can explain why qualia feel the way they do. The explanatory gap is closely related to the thought experiments that dominate the literature on qualia.

THE KNOWLEDGE ARGUMENT

One of these thought experiments is the case of Jackson's (1986) Mary, who is raised in a black and white environment in which she learns all the functional and physical facts about color vision. Nonetheless, 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 does she know 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 color vision, whether or not she is clever enough to figure it out on the basis of what she knows.

LEWIS denies that Mary acquires any new knowledge-that, insisting that she only acquires knowledge-how, abilities to imagine and recognize. But as Loar points out, the knowledge she acquires can appear in embedded contexts. For example, she may reason that if this is what it is

like to see red, then this is similar to what it is like to see orange. Lewis' ability analysis of Mary's knowledge has the same problem here that non-cognitive analyses of ethical language have in explaining the logical behavior of ethical predicates.

Here is a different (and in my view more successful) objection to Jackson (Horgan, 1984b; Peacocke, 1989; Loar, 1990; Papineau, 1993; van Gulick, 1993): What Mary acquires when she sees red is a new phenomenal concept, a recognitional disposition that allows her to pick out a certain type of phenomenal feel. 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. 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 physicalistic doctrines. Incidentally, the recognitional disposition account indicates how qualia could turn out to be relational; perhaps the recognitional disposition picks out a relational physical state of the brain or even a functional state. But see the criticism of Loar in FUNCTIONALISM.)

ABSENT QUALIA

Another familiar conundrum is the absent qualia hypothesis. If human beings can be described computationally, as is assumed by the research program 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 Shoemaker, 1975, 1981, and White, 1986.) Some thought experiments have appealed to oddball realizations of our functional organization, e.g. by 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 (1987) responds by making two additions to functionalism as spelled out in FUNCTIONALISM. The additions are designed to rule out oddball realizations of our functional organization of the ilk of the aforementioned economy. He suggests thinking of the functional roles in teleological terms and thinking of these roles as involving the details of human physiology. Economies don't have the states with the right sort of evolutionary "purpose", and their states are not physiological. On the first move, see FUNCTIONALISM. On the second, note that including physiology in our functional definitions of mental states will make them so specific to humans that they won't apply to other creatures that have mental states. Further, this idea violates the spirit of the functionalist proposal, which, being based on the computer analogy, abstracts from hardware realization. Functionalism without multiple hardware realizations is functionalism in name only.

THE INVERTED SPECTRUM

One familiar conundrum that uses a physicalistic idea of qualia against functionalist and intentionalist ideas is the famous inverted spectrum hypothesis, the hypothesis that things we both call "red" look to you the way things we both call "green" look to me, even though we are functionally (and therefore behaviorally) identical. A first step in motivating the inverted spectrum hypothesis is the possibility that the brain state that I have when I see red things is the same as the brain state that you have when you see green things, and conversely. (Nida-Rumelin, forthcoming, presents evidence that this is a naturally occurring phenomenon.) Therefor, it might be said, our

experiences are inverted. What is assumed here is a supervenience doctrine, that the qualitative content of a state supervenes on physiological properties of the brain.

There is a natural functionalist 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 can't be the same, since mine causes me to say "Its red", and to classify what I'm seeing as the same color as blood and fire hydrants, whereas yours causes you to say "Its 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 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 identified with brain states quite so "localized" as X-oscillations or Y-oscillations, but rather with more holistic brain states that include tendencies to classify objects together as the same color. 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. (Harman, 1990). So the fact that red things give me X-oscillations but they give you Y-oscillations doesn't 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 intra-personally, keeping the larger brain state that specifies the roles of X-oscillations in classifying things constant. He can insist on typing brain states for inter-personal 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", the fact that there is nothing known about the brain that can adequately explain the facts of experience. (See CONSCIOUSNESS.) 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 intra-personal inverted spectrum example. Think of this as a four stage process. (1) You have normal color vision. (2) You have color 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, colorwise, 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. 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. Second, the four-stage setup forces the opponents say what stage is the one where my description goes wrong. (See Shoemaker, 1981, Block, 1990.) Rey (1993) attacks (3), Dennett (1991) attacks (2) and (3), and White (1993) attacks (4). In my view, the most vulnerable stage is (3) because the functionalist can raise doubts about whether what its like to see red things *could* remain the same during the changes in responses that have to go on in the process of adaptation.

Why, an opponent might ask, is the inverted qualia argument against functionalism any more powerful than the 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. But though physical duplicates with inverted spectra may be imaginable, they are ruled out by a highly plausible principle that any materialist should accept: that qualia supervene on physical constitution. The thought experiments that I have been going through argue that even materialists should accept the possiblity of an inverted spectrum, and further, that for all we know, such cases are feasible via robotics or genetic engineering or even actual. And in so doing, they make the case for conceptual possibility stronger, for one is surer that something is genuinely conceptually possible if one can see how one might go about making it actual.

INVERTED EARTH

An interesting variant of the inverted spectrum thought-experiment is Inverted Earth (Block, 1990). Inverted Earth is a planet that differs from Earth in two relevant ways. First, everything is the complementary color of the corresponding earth object. The sky is yellow, the grass-like 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 color 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 inverters are inserted behind your eyes (and your body pigments are changed), you will notice no difference if you are placed in the bed of your counterpart on Inverted Earth. (Let's assume that the victim does not know anything about the science of color.)

Now consider the comparison between you and your counterpart on Inverted Earth. The counterpart could be your identical twin who was fitted with inverting lenses at birth and put up for adoption on Inverted Earth, or the counterpart could be you after you've been switched with your twin and have been living there for a long while. Looking at blue things give you Z-oscillations in the brain, yellow things give you W-oscillations; your twin gets the opposite. Now notice the interesting difference between this twin case and the one mentioned earlier: there can be perfect inversion in the *holistic* brain states as well as the local ones. At this moment, you both are looking at your respective skies. You get Z-oscillations because your sky is blue, he gets Z-oscillations because his sky is yellow. Your Z-oscillations make you say "How blue!", and his Z-oscillations make him say "How blue!" too. Indeed, we can take your brains to be molecular duplicates of one another. Then the principle of the supervenience of qualia on holistic brain state dictates that experientially, at the moment of looking at the skies, you and your twin have the same qualia.

But though you and your twin have the same qualia, you are functionally and intentionally inverted. If you are asked to match the color of the sky with a Munsell color chip, you will pick a blue one, but if your twin is shown the same (earth-made) Munsell chips, he will pick a yellow one. Further, when he says "How blue!" he *means* "How yellow!" Recall that the Inverted Earth dialect, of which he is a loyal member, has color words whose meanings are inverted with respect to ours. You and your twin are at that moment functionally and intentionally inverted, but qualitatively identical. So we have the converse of the inverted spectrum. And there is no problem about local

vs. holistic brain states as in the inter-subjective inverted spectrum; and no problem about whether qualia could persist unchanged through adaptation as in the intra-subjective inverted spectrum.

The argument that you and your twin are qualitatively the same can work either of two ways. We can assume the principle of supervenience of qualia on the brain, building the brain-identity of the twins into the story. Or we can run the story in terms of you being kidnapped, drugged, and placed in your twin's niche on Inverted Earth. What justifies the idea that your qualia are the same is that you notice no difference when you wake up in your Twin's bed after the switch; not appeal to supervenience is required.

Notice that the functional differences between these qualia-identical twins are long-arm functional differences (see FUNCTIONALISM) and the intentional differences are external intentional differences. Perhaps, you might say, the twins are not inverted in short-arm functional roles and narrow intentional content. The cure for this idea is to ask the question of what the purely internal functional or intentional differences could be that would define the difference between an experience as of red and an experience as of green. The natural answer would be to appeal to the internal aspects of beliefs and desires. We believe, for example, that blood is red but not that it is green. However, someone could have color experience despite having no standing beliefs or desires that differentiated colors. Imagine a person raised in a room where the color of everything is controlled by a computer, and nothing retains its color for more than 10 seconds. Or imagine a person whose color perception is normal but who has forgotten all color-facts.

There is no shortage of objections to these lines of reasoning. I will very briefly mention two closely related objections. It has been objected (Hardin, 1988) that red is intrinsically warm, whereas green is intrinsically cool, and thus inversion will either violate functional identity or yield an incoherent cool-red state. (Note, incidentally, that this isn't an objection to the inverted earth thought experiment, since that is a case of qualitative identity and functional *difference*--no functional identity is involved.) But the natural reply (Block, 1990) is that warm and cool can be inverted too. So long as there is no intrinsic connection between color qualia and behavior, the inverted spectrum is safe. But is there such an intrinsic connection? Dennett (1991) says there is. Blue calms, red excites. But perhaps this is due to culture and experience; perhaps people with very different cultures and experiences would have color experiences without this asymmetry. The research on this topic is equivocal; Dennett's sole reference, Humphrey (1992), describes it as "relatively second-rate", and that is also my impression. The fact that we don't know is itself interesting, however, for what it shows is that this asymmetry is no part of our color concepts. As Shoemaker (1981) points out, even if human color experience is genetically asymmetrical, there could nonetheless be people much like us whose color experience is not asymmetrical. So an inversion of the sort mentioned in the thought experiments is conceptually possible, even if it is not possible for the human species. But then color inversion may be possible for a closely related species whose color qualia are not in doubt, one which could perhaps be produced by genetic engineering. Functionalism would not be a very palatable doctrine if it were said to apply to some people's color experiences but not to others.

THE EXPLANATORY GAP AGAIN

At the outset, I mentioned the "explanatory gap", the idea that nothing now known about the brain, nor anything anyone has been able to imagine finding out would explain qualia. We can

distinguish inflationary and deflationary attitudes towards this gap among those who agree that the gap is unclosable. McGinn (1991) argues that the gap is unclosable because the fundamental nature of consciousness is inaccessible to us, though it might be accessible to creatures with very different sorts of minds. But a number of authors have favored a deflationary approach, arguing that the unclosability of the explanatory gap has to do with our concepts, not with nature itself. Horgan (1984), Levine (1993), Jackson (1993), Chalmers (1993) (and interestingly, McGinn, 1991 too) have contributed to working out the idea that reductive explanation in science depends on a priori analyses of the phenomena to be explained, usually in functional terms. (A version of this point was made in Nagel, 1974.) Consider Chalmers' example of the reductive explanation of life. Life can be roughly analyzed in terms of such general notions as metabolism and adaptation, or perhaps more specific notions such as digestion, reproduction and locomotion, and these concepts can themselves be given a functional analysis. Once we have explained these functions, suppose someone says "Oh yes, I see the explanation of those functions, but what about explaining *life*? We can answer that, a priori, to explain these functions *is* to explain life itself.

In some cases, the a priori analysis of the item to be explained is more complicated. Consider water. We can't give an a priori analysis of water as the colorless, odorless liquid in rivers and lakes called `water', because water might not have been colorless, it might have been called `glue', there might not have been lakes, etc. But we can formulate an a priori reference fixing definition of the sort that Kripke has emphasized: water = R(the colorless, odorless liquid in rivers and lakes called `water'), where the `R' is a rigidification operator that turns a definite description into a rigid designator. (A rigid designator picks out the same thing in all possible worlds in which the thing exists; for example, `Aristotle' is rigid. To rigidify a definite description is to treat it as a name for whatever the definite description *actually* picks out.) Thus, suppose we want to explain the fact that water dissolves salt It suffices to explain that H_2O dissolves salt and that H_2O is the colorless, odorless liquid in rivers and lakes called `water'. If someone objects that we have only explained how something colorless, odorless, etc. dissolves salt, not that water does, we can point out that it is a priori that water is the actual colorless, odorless, etc. substance. And if someone objects that we have only explained how H_2O dissolves salt, not how water does, we can answer that from the fact that H_2O is the colorless, odorless, etc. stuff and that, a priori, water is the (actual) colorless, odorless, etc., stuff, we can derive that water *is* H_2O .

The upshot is that closing the explanatory gap requires an a priori functional analysis of qualia. If Kripke (1980) is right that we pick out qualia by their qualitative character and not by their functional role, then no a priori reference fixing definition can be given for qualitative concepts of the sort that can be given for `water' and `life'. (Of course, if there is a true functional analysis that picks out a quale, it can be rigidified, but it still won't be an a priori characterization. Pain = R(Aunt Irma's favorite sensation) can be true and necessary without being a priori. And if the arguments about qualia inversion just sketched are right, there is no a priori conceptual analysis of qualitative concepts either, and so the explanatory gap is unclosable. As Chalmers points out, with a physical or a functional account, we can explain the functions associated with qualia, the capacity to classify things as red, for example. But once we have explained these functions, there will be a further question: why are these functions accompanied by qualia? Such a further question does not arise in the case of life and water precisely because of the availability of an a priori functional analysis.

It would be natural to suppose that the explanatory gap derives from the fact that neuroscientists have not yet come up with the required concepts to explain qualia. Nagel (1974) gives an analogy that suggests this idea. We are in the situation, he suggests, of a cave-man who is told that matter is energy. But he does not have the concepts to appreciate how this could be so. These concepts, however, are ones that some of us do have now, and it is a natural thought that a few hundred years from now, the concepts might be available to explain qualia physically. But the deflationary account of reductive explanation denies this, blaming the explanatory gap on our ordinary concepts, not on science.

Bibliography

Block, N. (1990) Inverted earth. In *Philosophical Perspectives* 4 ed J. Tomberlin. Ridgeview

Chalmers, D.J. (1993) *Toward a Theory of Consciousness*. University of Indiana Ph.D. thesis

Davies, M. & Humphreys, G. (1993a) *Consciousness*. Blackwell:Oxford

Dennett, D. (1988) `Quining Qualia.' In A. Marcel & E. Bisiach (eds) *Consciousness in Contemporary Society*. Oxford University Press: Oxford
(1991) *Consciousness Explained*. Little Brown: New York

Flanagan, O. (1992) *Consciousness Reconsidered* MIT Press

Hardin, C. (1988) *Color for Philosophers* Hackett: Indianapolis

Harman, G. (1990) `The intrinsic quality of experience.' In *Philosophical Perspectives* 4 ed J. Tomberlin. Ridgeview.

Horgan, T. (1984a) `Supervenience and cosmic hermeneutics'. *Southern Journal of Philosophy Supplement* 22: 19-38
(1984b) `Jackson on physical information and qualia'. *Philosophical Quarterly* 34

Jackson, F. (1986) `What Mary didn't know.' *Journal of Philosophy* 83: 291-95

Jackson, F. (1993) `Armchair metaphysics'. In J. O'Leary-Hawthorne and M. Michael (eds) *Philosophy in Mind*. Kluwer

Kripke, S. (1980) *Naming and Necessity* Harvard University Press:Cambridge

Levine, J. (1993) `On leaving out what it is like.' In Davies and Humphreys (1993a)

Loar, B. (1990) `Phenomenal properties.' In J. Tomberlin (ed) *Philosophical Perspectives: Action Theory and Philosophy of Mind*. Ridgeview.

Lycan, W. (1987) *Consciousness* MIT Press: Cambridge

McGinn, C. (1991) *The Problem of Consciousness*. Blackwell

Nida-Rumelin, M. (forthcoming) `Pseudonormal vision. An actual case of qualia inversion?' In *Philosophical Studies*

Papineau, D. (1993) `Physicalism, Consciousness and the Antipathetic Fallacy'. *The Australasian Journal of Philosophy* 71, 2: 169-184

Peacocke, C. (1989) `No resting place: a critical notice of The View from Nowhere', *The Philosophical Review* 98, 65-82.

Rey, G. (1993) `Sensational Sentences Switched'. *Philosophical Studies* 70, 1:

Shoemaker, S. (1975) `Functionalism and qualia.' *Philosophical Studies* 27: 291-315.
(1981) `Absent qualia are impossible--a reply to Block'. *The Philosophical Review* 90,4:581-599

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

White, Stephen L. (1986): `Curse of the qualia', *Synthese* 68: 333-368.
(1993) `Color and the narrow contents of experience' Paper delivered at the Eastern Division of the American Philosophical Association.