

Epiphenomenal Qualia

FRANK JACKSON

Frank Jackson is professor of philosophy at Australian National University. He has made influential contributions to several areas of philosophy and is the author of a widely discussed critique of physicalism. His books include *Perception: A Representative Theory* and *From Metaphysics to Ethics: A Defense of Conceptual Analysis*.

Reprinted from *The Philosophical Quarterly*, 32, by permission of the journal.

It is undeniable that the physical, chemical and biological sciences have provided a great deal of information about the world we live in and about ourselves. I will use the label "physical information" for this kind of information, and also for information that automatically comes along with it. For example, if a medical scientist tells me enough about the processes that go on in my nervous system, and about how they relate to happenings in the world around me, to what has happened in the past and is likely to happen in the future, to what happens to other similar and dissimilar organisms, and the like, he or she tells me--if I am clever enough to fit it together appropriately--about what is often called the functional role of those states in me (and in organisms in general in similar cases). This information, and its kin, I also label "physical."

I do not mean these sketchy remarks to constitute a definition of "physical information," and of the correlative notions of physical property, process, and so on, but to indicate what I have in mind here. It is well known that there are problems with giving a precise definition of these notions, and so of the thesis of Physicalism that all (correct) information is physical information. But--unlike some--I take the question of definition to cut across the central problems I want to discuss in this paper.

I am what is sometimes known as a "qualia freak." I think that there are certain features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes. Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a rose, hearing a loud noise or seeing the sky.

There are many qualia freaks, and some of them say that their rejection of Physicalism is an unargued intuition. I think that they are being unfair to themselves. They have the following argument. Nothing you could tell of a physical sort captures the smell of a rose, for instance. Therefore, Physicalism is false. By our lights this is a perfectly good argument. It is obviously not to the point to question its validity, and the premise is intuitively obviously true both to them and to me.

I must, however, admit that it is weak from a polemical point of view. There are, unfortunately for us, many who do not find the premise intuitively obvious. The task then is to present an argument whose premises are obvious to all, or at least to as many as possible. This I try to do in Section I with what I will call "the Knowledge argument." . . . In Section IV I tackle the question 'of the causal role of qualia. The major factor in stopping people from admitting qualia is the belief that they would have to be given a causal role with respect to the physical world and especially the brain; and it is hard to do this without sounding like someone who believes in fairies. I seek in Section IV to turn this objection by arguing that the view that qualia are epiphenomenal is a perfectly possible one.

I. THE KNOWLEDGE ARGUMENT FOR QUALIA

People vary considerably in their ability to discriminate colours. Suppose that in an experiment to catalogue this variation Fred is discovered. Fred has better colour vision than anyone else on record; he makes every discrimination that anyone has ever made, and moreover he makes one that we cannot even begin to make. Show him a batch of ripe tomatoes and he sorts them into two

roughly equal groups and does so with complete consistency. That is, if you blindfold him, shuffle the tomatoes up, and then remove the blindfold and ask him to sort them out again, he sorts them into exactly the same two groups.

We ask Fred how he does it. He explains that all ripe tomatoes do not look the same colour to him, and in fact that this is true of a great many objects that we classify together as red. He sees two colours where we see one, and he has in consequence developed for his own use two words "red1 " and "red2" to mark the difference. Perhaps he tells us that he has often tried to teach the difference between red1 and red2 to his friends but has got nowhere and has concluded that the rest of the world is red1-red2 colour-blind--or perhaps he has had partial success with his children, it doesn't matter. In any case he explains to us that it would be quite wrong to think that because "red" appears in both "red1" and "red2" that the two colours are shades of the one colour. He only uses the common term "red" to fit more easily into our restricted usage. To him red1 and red2 are as different from each other and all the other colours as yellow is from blue. And his discriminatory behaviour bears this out: he sorts red1 from red2 tomatoes with the greatest of ease in a wide variety of viewing circumstances. Moreover, an investigation of the physiological basis of Fred's exceptional ability reveals that Fred's optical system is able to separate out two groups of wavelengths in the red spectrum as sharply as we are able to sort out yellow from blue.

I think that we should admit that Fred can see, really see, at least one more colour than we can; red2 is a different colour from red1. We are to Fred as a totally red-green colour-blind person is to us. H. G. Wells' story "The Country of the Blind" is about a sighted person in a totally blind community. This person never manages to convince them that he can see, that he has an extra sense. They ridicule this sense as quite inconceivable, and treat his capacity to avoid falling into ditches, to win fights and so on as precisely that capacity and nothing more. We would be making their mistake if we refused to allow that Fred can see one more colour than we can.

What kind of experience does Fred have when he sees red1 and red2? What is the new colour or colours like? We would dearly like to know but do not; and it seems that no amount of physical information about Fred's brain and optical system tells us. We find out perhaps that Fred's cones respond differentially to certain light waves in the red section of the spectrum that make no difference to ours (or perhaps he has an extra cone) and that this leads in Fred to a wider range of those brain states responsible for visual discriminatory behaviour. But none of this tells us what we really want to know about his colour experience. There is something about it we don't know. But we know, we may suppose, everything about Fred's body, his behaviour and dispositions to behaviour and about his internal physiology, and everything about his history and relation to others that can be given in physical accounts of persons. We have all the physical information. Therefore, knowing all this is not knowing everything about Fred. It follows that Physicalism leaves something out.

To reinforce this conclusion, imagine that as a result of our investigations into the internal workings of Fred we find out how to make everyone's physiology like Fred's in the relevant respects; or perhaps Fred donates his body to science and on his death we are able to transplant his optical system into someone else--again the fine detail doesn't matter. The important point is that such a happening would create enormous interest. People would say, "At last we will know what it is like to see the extra colour, at last we will know how Fred has differed from us in the way he has struggled to tell us about for so long". Then it cannot be that we knew all along all about Fred. But *ex hypothesi* we did know all along everything about Fred that features in the physicalist scheme; hence the physicalist scheme leaves something out.

Put it this way. *After* the operation, we will know *more* about Fred and especially about his colour experiences. But beforehand we had all the physical information we could desire about his body and brain, and indeed everything that has ever featured in physicalist accounts of mind and consciousness. Hence there is more to know than all that. Hence Physicalism is incomplete. Fred and the new colour(s) are of course essentially rhetorical devices. The same point can be made with normal people and familiar colours. Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like "red," "blue," and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air

from the lungs that results in the uttering of the sentence "The sky is blue." (It can hardly be denied that it is in principle possible to obtain all this physical information from black and white television, otherwise the Open University would of necessity need to use colour television.)

What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.

Clearly the same style of Knowledge argument could be deployed for taste, hearing, the bodily sensations and generally speaking for the various mental states which are said to have (as it is variously put) raw feels, phenomenal features or qualia. The conclusion in each case is that the qualia are left out of the physicalist story. And the polemical strength of the Knowledge argument is that it is so hard to deny the central claim that one can have all the physical information without having all the information there is to have. . . .
{Sections II and III of this essay are not included in this excerpt.}

IV. THE BOGEY OF EPIPHENOMENALISM

Is there any really good reason for refusing to countenance the idea that qualia are causally impotent with respect to the physical world? I will argue for the answer no, but in doing this I will say nothing about two views associated with the classical epiphenomenalist position. The first is that mental states are inefficacious with respect to the physical world. All I will be concerned to defend is that it is possible to hold that certain properties of certain mental states, namely those I've called qualia, are such that their possession or absence makes no difference to the physical world. The second is that the mental is totally causally inefficacious. For all I will say it may be that you have to hold that the instantiation of *qualia* makes a difference to *other mental states* though not to anything physical. Indeed general considerations to do with how you could come to be aware of the instantiation of qualia suggest such a position.

Three reasons are standardly given for holding that a quale like the hurtfulness of a pain must be causally efficacious in the physical world, and so, for instance, that its instantiation must sometimes make a difference to what happens in the brain. None, I will argue, has any real force. (I am much indebted to Alec Hyslop and John Lucas for convincing me of this.)

(i) It is supposed to be just obvious that the hurtfulness of pain is partly responsible for the subject seeking to avoid pain, saying "It hurts" and so on. But, to reverse Hume, anything can fail to cause anything. No matter how often B follows A, and no matter how initially obvious the causality of the connection seems, the hypothesis that A causes B can be overturned by an over-arching theory which shows the two as distinct effects of a common underlying causal process.

To the untutored the image on the screen of Lee Marvin's fist moving from left to right immediately followed by the image of John Wayne's head moving in the same general direction looks as causal as anything. And of course throughout countless Westerns images similar to the first are followed by images similar to the second. All this counts for precisely nothing when we know the over-arching theory concerning how the relevant images are both effects of an underlying causal process involving the projector and the film. The epiphenomenalist can say exactly the same about the connection between, for example, hurtfulness and behaviour. It is simply a consequence of the fact that certain happenings in the brain cause both.

(ii) The second objection relates to Darwin's Theory of Evolution. According to natural selection the traits that evolve over time are those conducive to physical survival. We may assume that qualia evolved over time—we have them, the earliest forms of life do not—and so we should expect qualia to be conducive to survival. The objection is that they could hardly help us to survive if they do nothing to the physical world.

The appeal of this argument is undeniable, but there is a good reply to it. Polar bears have particularly thick, warm coats. The Theory of Evolution explains this (we suppose) by pointing out that having a thick, warm coat is conducive to survival in the Arctic. But having a thick coat goes along with having a heavy coat, and having a heavy coat is not conducive to survival. It slows the animal down.

Does this mean that we have refuted Darwin because we have found an evolved trait--having a heavy coat--which is not conducive to survival? Clearly not. Having a heavy coat is an unavoidable concomitant of having a warm coat (in the context, modern insulation was not available), and the advantages for survival of having a warm coat outweighed the disadvantages of having a heavy one. The point is that all we can extract from Darwin's theory is that we should expect any evolved characteristic to be either conducive to survival or a by-product of one that is so conducive. The epiphenomenalist holds that qualia fall into the latter category. They are a byproduct of certain brain processes that are highly conducive to survival.

(iii) The third objection is based on a point about how we come to know about other minds. We know about other minds by knowing about other behaviour, at least in part. The nature of the inference is a matter of some controversy, but it is not a matter of controversy that it proceeds from behaviour. That is why we think that stones do not feel and dogs do feel. But, runs the objection, how can a person's behaviour provide any reason for believing he has qualia like mine, or indeed any qualia at all, unless this behaviour can be regarded as the outcome of the qualia. Man Friday's footprint was evidence of Man Friday because footprints are causal outcomes of feet attached to people. And an epiphenomenalist cannot regard behaviour, or indeed anything physical, as an outcome of qualia.

But consider my reading in *The Times* that Spurs won. This provides excellent evidence that *The Telegraph* has also reported that Spurs won, despite the fact that (I trust) *The Telegraph* does not get the results from *The Times*. They each send their own reporters to the game. The *Telegraph's* report is in no sense an outcome of *The Times'*, but the latter provides good evidence for the former nevertheless.

The reasoning involved can be reconstructed thus. I read in *The Times* that Spurs won. This gives me reason to think that Spurs won because I know that Spurs' winning is the most likely candidate to be what caused the report in *The Times*. But I also know that Spurs' winning would have had many effects, including almost certainly a report in *The Telegraph*. I am arguing from one effect back to its cause and out again to another effect. The fact that neither effect causes the other is irrelevant. Now the epiphenomenalist allows that qualia are effects of what goes on in the brain. Qualia cause nothing physical but are caused by something physical. Hence the epiphenomenalist can argue from the behaviour of others to the qualia of others by arguing from the behaviour of others back to its causes in the brains of others and out again to their qualia.

You may well feel for one reason or another that this is a more dubious chain of reasoning than its model in the case of newspaper reports. You are right. The problem of other minds is a major philosophical problem, the problem of other newspaper reports is not. But there is no special problem of epiphenomenalism as opposed to, say, Interactionism here.

There is a very understandable response to the three replies I have just made. "All right, there is no knockdown refutation of the existence of epiphenomenal qualia. But the fact remains that they are an excrescence. They do nothing, they explain nothing, they serve merely to soothe the intuitions of dualists, and it is left a total mystery how they fit into the world view of science. In short we do not and cannot understand the how and why of them."

This is perfectly true; but is no objection to qualia, for it rests on an overly optimistic view of the human animal, and its powers. We are the products of Evolution. We understand and sense what we need to understand and sense in order to survive. Epiphenomenal qualia are totally irrelevant to survival. At no stage of our evolution did natural selection favour those who could make sense of how they are caused and the laws governing them, or in fact why they exist at all. And that is why we can't.

It is not sufficiently appreciated that Physicalism is an extremely optimistic view of our powers. If it is true, we have, in very broad outline admittedly, a grasp of our place in the scheme of things. Certain matters of sheer complexity defeat us--there are an awful lot of neurons--but in principle we have it all. But consider the antecedent probability that everything in the Universe be of a kind that is relevant in some way or other to the survival of homo sapiens. It is very low surely. But then one must admit that it is very likely that there is a part of the whole scheme of things, maybe a big part, which no amount of evolution will ever bring us near to knowledge about or understanding. For the simple reason that such knowledge and understanding are irrelevant to survival.

Physicalists typically emphasise that we are a part of nature on their view, which is fair enough. But if we are a part of nature, we are as nature has left us after however many years of evolution it is, and each step in that evolutionary progression has been a matter of chance constrained just by the need to preserve or increase survival value. The wonder is that we understand as much as we do, and there is no wonder that there should be matters which fall quite outside our comprehension. Perhaps exactly how epiphenomenal qualia fit into the scheme of things is one such.

This may seem an unduly pessimistic view of our capacity to articulate a truly comprehensive picture of our world and our place in it. But suppose we discovered living on the bottom of the deepest oceans a sort of sea slug which manifested intelligence. Perhaps survival in the conditions required rational powers. Despite their intelligence, these sea slugs have only a very restricted conception of the world by comparison with ours, the explanation for this being the nature of their immediate environment. Nevertheless they have developed sciences which work surprisingly well in these restricted terms. They also have philosophers, called slugists. Some call themselves tough-minded slugists, others confess to being soft-minded slugists. The tough-minded slugists hold that the restricted terms (or ones pretty like them which may be introduced as their sciences progress) suffice in principle to describe everything without remainder. These tough-minded slugists admit in moments of weakness to a feeling that their theory leaves something out. They resist this feeling and their opponents, the soft-minded slugists, by pointing out--absolutely correctly--that no slugist has ever succeeded in spelling out how this mysterious residue fits into the highly successful view that their sciences have and are developing of how their world works. Our sea slugs don't exist, but they might. And there might also exist super beings which stand to us as we stand to the sea slugs. We cannot adopt the perspective of these super beings, because we are not them, but the possibility of such a perspective is, I think, an antidote to excessive optimism.