

Jackson, Frank. (1982) "Epiphenomenal Qualia [excerpt]." [+ passages from "What Mary Didn't Know" (1986) *Journal of Philosophy* LXXXIII 5: 291-95.]. *Philosophical Quarterly* 32: 127-36.

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 happening 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.

[Physicalism is not the noncontroversial thesis that the actual world is largely physical, but the challenging thesis that it is entirely physical. This is why physicalists must hold that complete physical knowledge is complete knowledge simpliciter. For suppose it is not complete: then our world must differ from a world, $W(P)$, for which it is complete, and the difference must be in nonphysical facts: for our world and $W(P)$ agree in all matters physical. Hence, physicalism would be false at our world [though contingently so, for it would be true at $W(P)$].¹ (Jackson 1986)

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.

¹ "The claim here is not that, if physicalism is true, only what is expressed in explicitly physical language is an item of knowledge. It is that, if physicalism is true, then if you know everything expressed or expressible in explicitly physical language, you know everything." (Jackson 1986, footnote 1)

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 II I contrast the Knowledge argument with the Modal argument², and in section III with the "What is it like to be" argument. . . .

I. The Knowledge argument for qualia

People vary considerably in their ability to discriminate colors. Suppose that in an experiment to catalog this variation, Fred is discovered. Fred has better color 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 color to him, and in fact that this is true of a great many objects that we classify together as red. He sees two colors 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 color-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 colors are shades of the one color. 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 colors as yellow is from blue. And his discriminatory behavior 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 color than we can; red₁ is a different color from red₂. We are to Fred as a totally red-green color-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 color than we can.

² We will not read Section II, but if you are interested, feel free to read about "The Zombie Argument for Dualism" in Chapter 7 of Montero's *On the Philosophy of Mind*.

What kind of experience does Fred have when he sees red₁ and red₂? What is the new color or colors 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 behavior. But none of this tells us what we really want to know about his color experience. There is something about it we don't know. But we know, we may suppose, everything about Fred's body, his behavior and dispositions to behavior 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 color, 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*³ 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 color 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 color(s) are of course essentially rhetorical devices. The same point can be made with normal people and familiar colors. Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room, [educated through black-and-white books and through lectures relayed on black-and-white television] . . . monitor. She specializes 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 wavelength 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 color television.) [In this way she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of 'physical' which includes everything in completed physics, chemistry, and neurophysiology, and all

³ according to the hypothesis proposed

there is to know about the causal and relational facts consequent upon all this, including of course functional roles. If physicalism is true, she knows all there is to know. For to suppose otherwise is to suppose that there is more to know than every physical fact, and that is what physicalism denies.] (Jackson 1986)

What will happen when Mary is released from her black and white room or is given a color 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. [She will learn what it is like to see something red, say. This is rightly described as learning--she will not say "ho, hum."] But then it is inescapable that her previous knowledge was incomplete[:Mary [did] not know all there is to know]. But she had *all* the physical information. *Ergo* there is more [knowledge] to have than that, and physicalism is false. [This is the knowledge argument against physicalism in one of its manifestations.

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. . . .

III. The "What is it like to be" argument

In "What is it like to be a bat?" Thomas Nagel argues that no amount of physical information can tell us what it is like to be a bat, and indeed that we, human beings, cannot imagine what it is like to be a bat."¹⁰ His reason is that what this is like can only be understood from a bat's point of view, which is not our point of view, and is not something capturable in physical terms, which are essentially terms understandable equally from many points of view.

It is important to distinguish this argument from the Knowledge argument. When I complained that all the physical knowledge about Fred was not enough to tell us what his special color experience was like, I was not complaining that we weren't finding out what it is like to *be* Fred. I was complaining that there is something *about* his experience, a property of it, of which we were left ignorant. And if and when we come to know what this property is we still will not know what it is like to *be* Fred, but we will know more *about* him. No amount of knowledge about Fred, be it physical or not, amounts to knowledge "from the inside" considering Fred. We are not Fred. There is thus a whole set of items of knowledge expressed by forms of words like 'that is *I myself* who is ...' which Fred has and we simply cannot have because we are not him.

When Fred sees the color he alone can see, one thing he knows is the way his experience of it differs from his experience of seeing red and so on; *another* is that he himself is seeing it. Physicalist and qualia freaks alike should acknowledge that no amount of information of whatever kind that *others* have *about* Fred amounts to knowledge of the second. My complaint though concerned the first and was that the special quality of his

experience is certainly a fact about it and one which physicalism leaves out because no amount of physical information told us what it is.

Nagel speaks as if the problem he is raising is one of extrapolating from knowledge of one experience to another, of imagining what an unfamiliar experience would be like on the basis of familiar ones. In terms of Hume's example, from knowledge of some shades of blue we can work out what it would be like to see other shades of blue.⁴ Nagel argues that the trouble with bats et al. is that they are too unlike us. It is hard to see an objection to physicalism here. Physicalism makes no special claims about the imaginative or extrapolative powers of human beings, and it is hard to see why it need do so.

Anyway, our Knowledge argument makes no assumptions on this point. If physicalism were true, enough physical information about Fred would obviate any need to extrapolate or to perform special feats of imagination or understanding in order to know all about his special color experience. *The information would already be in our possession.* But it clearly isn't. That was the nub of the argument. . . .

"What Mary Didn't Know" [excerpt]

Three Clarifications

The knowledge argument does not rest on the dubious claim that logically you cannot imagine what sensing red is like unless you have sensed red. Powers of imagination are not to the point. The contention about Mary is not that, despite her fantastic grasp of neurophysiology and everything else physical, she *could not imagine* what it is like to sense red; it is that, as a matter of fact, she *would not know*. But if physicalism is true, she would know; and no great powers of imagination would be called for. Imagination is a faculty that those who *lack* knowledge need to fall back on.

Second, the intensionality of knowledge⁵ is not to the point. The argument does not rest on assuming falsely that, if *S* knows that *a* is *F* and if $a = b$, then *S* knows that *b* is *F*. It is concerned with the nature of Mary's total body of knowledge before she is released: is it complete, or do some facts escape it? What is to the point is that *S* may know that *a* is *F* and *know* that $a = b$, yet arguably not know that *b* is *F*, by virtue of not being sufficiently logically alert to follow the consequences through. If Mary's lack of knowledge were at all like this, there would be no threat to physicalism in it. But it is very hard to believe that her lack of knowledge could be remedied merely by her explicitly following through enough logical consequences of her vast physical knowledge. Endowing her with great logical acumen and persistence is not in itself enough to fill in the gaps in her knowledge. On being let out [of the black-and-white room], she will not say "I could have worked all this out before by making some more purely logical inferences."

⁴ Remember the passage we read from Hume's *An Enquiry Concerning Human Understanding*, where he considers a complication to his empiricist theory of knowledge. Our purported ability to imagine what the "missing shade of blue" is like, intermediate between two shades that we have actually seen and experienced, complicates his theory that we only know what we have learned through our senses or what we can combine from previously experienced sensations (e.g. "golden horse" = gold + horse).

⁵ i.e., the meaning of the term "knowledge"

Third, the knowledge Mary lacked which is of particular point for the knowledge argument against physicalism is *knowledge about the experiences of others*, not about her own. When she is let out, she has new experiences, color experiences she has never had before, It is not, therefore, an objection to physicalism that she learns *something* on being let out. Before she was let out, she could not have known facts about her experience of red, for there were no such facts to know. That physicalist and nonphysicalist alike can agree on. After she is let out, things change; and physicalism can happily admit that she learns this; after all, some physical things will change, for instance, her brain states and their functional roles. The trouble for physicalism is that, after Mary sees her first ripe tomato, she will realize how impoverished her conception of the mental life of *others* has been *all along*. She will realize that there was, all the time she was carrying out her laborious investigations into the neurophysiologies of others and into the functional roles of their internal states, something about these people she was quite unaware of. All along their experiences (or many of them, those got from tomatoes, the sky...) had a feature conspicuous to them but until now hidden from her (in fact, not in logic). But she knew all the physical facts about them all along; hence, what she did not know until her release is not a physical fact about their experiences. But it is a fact about them. That is the trouble for physicalism,

[Paul] Churchland's Three Objections

(i) Churchland's first objection is that the knowledge argument contains a defect that "is simplicity itself". The argument equivocates on the sense of 'knows about'. How so? Churchland suggests that the following is "a conveniently tightened version" of the knowledge argument:

- (1) Mary knows everything there is to know about brain states and their properties,
- (2) It is not the case that Mary knows everything there is to know about sensations and their properties.

Therefore, by Leibniz's law⁶,

- (3) Sensations and their properties \neq brain states and their properties.

Churchland observes, plausibly enough, that the type or kind of knowledge involved in premise 1 is distinct from the kind of knowledge involved in premise 2. We might follow his lead and tag the first 'knowledge by description', and the second 'knowledge by acquaintance'; but, whatever the tags, he is right that the displayed argument

⁶ Roughly, if there is at least one property that x has and y does not, then x and y are not identical, but rather distinct.

In this case, x = brain states and their properties, and y = sensations and their properties. Churchland claims that the knowledge argument against physicalism says that Mary knows x , but Mary does not know y . If being-known-by-Mary is a property that x has but y does not, then x and y cannot be identical. Read more about this principle here: <http://www.oberlin.edu/faculty/mwallace/LeibnizsLaw.html>

involves a highly dubious use of Leibniz's law⁷.

My reply is that the displayed argument may be convenient, but it is not accurate. It is not the knowledge argument. Take, for instance, premise 1. The whole thrust of the knowledge argument is that Mary (before her release) does *not* know everything there is to know about brain states and their properties, because she does not know about certain qualia associated with them. What is complete, according to the argument, is her knowledge of matters physical. A convenient and accurate way of displaying the argument is:

(1)' Mary (before her release) knows everything physical there is to know about other people.

(2)' Mary (before her release) does not know everything there is to know about other people (because she *learns* something about them on her release).

Therefore,

(3)' There are truths about other people (and herself) which escape the physicalist story.

What is immediately to the point is not the kind, manner, or type of knowledge Mary has, but *what* she knows. What she knows beforehand is *ex hypothesi* everything physical there is to know, but is it everything there is to know? That is the crucial question.

There is, though, a relevant challenge involving questions about kinds of knowledge. It concerns the *support* for premise 2'. The case for premise 2' is that Mary learns something on her release, she acquires knowledge, and that entails that her knowledge beforehand (*what* she knew, never mind whether by description, acquaintance, or whatever) was incomplete. The challenge, mounted by David Lewis and Laurence Nemirow, is that on her release Mary does *not* learn something or acquire knowledge in the relevant sense. What Mary acquires when she is released is a certain representational or imaginative ability; it is knowledge *how* rather than knowledge *that*⁸. Hence, a physicalist can admit that Mary acquires something very significant of a knowledge kind – which can hardly be denied – without admitting that this shows that her earlier factual knowledge is defective. She knew all *that* there was to know about the experiences of others beforehand, but lacked an ability until after her release.

Now it is certainly true that Mary will acquire abilities of various kinds after her release. She will, for instance, be able to imagine what seeing red is like, be able to remember

⁷ Dubious because it's up to dispute whether 'being known' is rightly described as a *property* of something like brain states or sensations. At least, it doesn't seem to be a property intrinsic to those things; rather, it would be a relational property between those things and Mary.

⁸ "Knowledge how" is what enables someone to do something, i.e., to perform a skill; "knowledge that", on the other hand, is understanding or comprehension of a fact. E.g., I have knowledge how to swim, but I have knowledge that I need to kick my feet continuously to move forward while swimming. The best-known description of this difference can be found in Gilbert Ryle's *The Concept of Mind* (1949). Ryle claims that there is a vast difference between knowledge how and knowledge that; other philosophers hold an opposing view that all knowledge how is just the sum of knowledge that for all the facts that add up to the ability to do something.

what it is like, and be able to understand why her friends regarded her as so deprived (something which, until her release, had always mystified her). But is it plausible that that is *all* she will acquire? Suppose she received a lecture on skepticism about other minds while she was incarcerated [i.e., stuck in the black-and-white room]. On her release she sees a ripe tomato in normal conditions, and so has a sensation of red. Her first reaction is to say that she now knows more about the kind of experiences others have when looking at ripe tomatoes. She then remembers the lecture and starts to worry. Does she really know more about what their experiences are like, or is she indulging in a wild generalization from one case? In the end she decides she does know, and that skepticism is mistaken (even if, like so many of us, she is not sure how to demonstrate its errors). What was she to-ing and fro-ing about – her abilities? Surely not; her representational abilities were a known constant throughout. What else then was she agonizing about than whether or not she had gained factual knowledge of others? There would be nothing to agonize about if ability was *all* she acquired on her release.

I grant that I have no *proof* that Mary acquires on her release, as well as abilities, factual knowledge about the experiences of others – and not just because I have no disproof of skepticism. My claim is that the knowledge argument is a valid argument from highly plausible, though admittedly not demonstrable, premises to the conclusion that physicalism is false. And that, after all, is about as good an objection as one could expect in this area of philosophy.

(ii) Churchland's second objection is that there must be something wrong with the argument, for it proves too much. Suppose Mary received a special series of lectures over her black-and-white television from a full-blown dualist, explaining the "laws" governing the behavior of "ectoplasm"⁹ and telling her about qualia. This would not affect the plausibility of the claim that on her release she learns something. So if the argument works against physicalism, it works against dualism too.

My reply is that lectures about qualia over black-and-white television do not tell Mary all there is to know about qualia. They may tell her some things about qualia, for instance, that they do not appear in the physicalist's story, and that the quale we use 'yellow' for is nearly as different from the one we use 'blue' for as is white from black. But why should it be supposed that they tell her everything about qualia? On the other hand, it is plausible that lectures over black-and-white television might in principle tell Mary everything in the physicalist's story. You do not need color television to learn physics or functionalist psychology. To obtain a good argument against dualism . . . , the premise in the knowledge argument that Mary has the full story according to physicalism before her release has to be replaced by a premise that she has the full story *according to dualism*. The former is plausible; the latter is not. Hence, there is no "parity of reasons" trouble for dualists who use the knowledge argument.

(iii) Churchland's third objection is that the knowledge argument claims "that Mary could not even *imagine* what the relevant experience would be like, despite her exhaustive neuroscientific knowledge, and hence must still be missing certain crucial

⁹ Some versions of dualism, which describe the mental and the physical as different types of substances, use the term "ectoplasm" to characterize the substance of mental entities.

information", a claim he goes on to argue against.

But, as we emphasized earlier, the knowledge argument claims that Mary would not *know* what the relevant experience is like. What she could imagine is another matter. If her knowledge is defective, despite being all there is to know according to physicalism, then physicalism is false, whatever her powers of imagination.