

Class #23 - Epiphenomenalism
Jackson, "Epiphenomenal Qualia"

I. Functionalism, Liberals and Chauvinists

We have looked at four theories of mind: dualism, behaviorism, identity theory, and functionalism. While functionalism seems promising, we also looked at two important criticisms of the view. The absent qualia argument and Searle's Chinese room example are both intended to show that functionalism is too liberal.

It ascribes mental states to artifacts and organisms which we lack good reasons to believe really have mental states: the homunculi-headed robot, the person in the Chinese room, the Chinese nation.

There are two kinds of functionalist responses to charges of excessive liberalism.

First, the functionalist can adjust the theory or our interpretation of the purported counter-examples so that functionalism no longer seems too liberal.

Second, the functionalist can accept the charge of liberalism, but say that liberalism is not a bad thing. I'll call these first-class and second-class responses, without intending to suggest that one response is better than the other.

First-class responses include providing reasons why the purported counterexamples, like the Chinese-nation brain, are not persons with minds despite their functional organization.

Second-class responses allow that such artificial persons, functional equivalents of persons, have mental states, but that they are not counter-examples to functionalism.

For a first-class response, we have to find a way to disallow the purported counter-examples from counting against the theory.

One way to do that is to appeal to a mature neuroscience to determine the boundaries of our psychological theory.

To make such a response, the functionalist could claim that the purported counterexamples lack minds because they don't have the appropriate neural networks.

Such a functionalism, sometimes called psychofunctionalism, looks a lot like identity theory.

Both identity theory and psychofunctionalism tie mentality to specifically human, physiological characteristics.

By appealing to particular hardware instantiations of mind, psychofunctionalism errs in the other direction.

It suffers from chauvinism, as the identity theory did.

The psychofunctionalist might try to avoid charges of chauvinism by claiming that there is a universal theory of psychology.

This suggestion, though, seems to lead at best to a disjunctive psychological theory: something has this mental state if it has this physical state, or that one, or this other one...

We have seen that disjunctive theories are not attractive.

If we combine this disjunctive theory with equipotentiality, we are pretty close to saying that any mental state can be correlated with any physical state.

That is a lot like saying nothing.

For a second-class response, the functionalist might deny that qualia are real in order to show the purported counterexamples are not relevantly dissimilar from us.

If we lack qualia, then we shouldn't be bothered by the apparent absence of qualia in the purported counterexamples.

It's not unprecedented to wonder whether the internal appearance of mental states is illusory.
The radical behaviorist already made this claim.
The claim may seem implausible, but perhaps we should reconsider it.
The functionalist can argue that qualia have no room in our best scientific psychological theory.
There may be nothing that it is like to be us.
If my phenomenal states aren't real, then there is nothing missing in the artificial minds which looked like counterexamples.
We could attribute intentional states to functional equivalents of persons, these purported counterexamples of Turing machines running theories of mind.
The homunculi-headed robots are ascribed mental states, but such ascriptions are consistent with what we know about psychology.

Note that the psychofunctionalist also denies the reality of qualia, as long as a mature neuroscience makes no reference to them.
And, it is difficult to see how mature neuroscience can make reference to qualia.
Mature scientific theories will make use of wavelengths of light, but not the way colors feel to us.
So, both the first-class response and the second-class response raise the question of the reality of my conscious experience.

Functionalism is the most promising theory of mind of the twentieth century.
It is the most successful theory compatible with physicalism.
But, it seems to have a problem accommodating first-personal accounts of qualitative states.
As Descartes observed long ago, sound is not the motion of air.
So, if a contemporary theory of mind is going to be satisfactory, it has to come to terms with mental states as they appear from the inside.
One option is to denigrate the reality of qualitative states.
Another option is to give up on physicalism.

II. The Denial of Qualia

The problem of consciousness seems to affect all kinds of physicalism: behaviorism, identity theory, and functionalism (when taken as token physicalism).
We might wonder whether any purely physicalistic theory of the mind can account for our conscious experience.

We have seen the problems with behaviorism and functionalism in Block's absent qualia argument, and in Searle's Chinese Room example.
Both theories appear to be too liberal since they attribute mental states to functional equivalents of minds.
Searle focused on whether machines could have intentions.
Block focused on qualia.
So, whether we look at qualitative states (qualia) or intentional states (beliefs, desires, etc.), consciousness seems to be an issue for the functionalist or behaviorist.

The claim that a scientific theory will omit elements of our conscious experience has a long history.
Einstein is said to have claimed that science could never give us the taste of soup.
Wittgenstein said that nothing would serve as well as something about which nothing could be said.
Daniel Dennett denies even that qualia are something about which nothing can be said.
He imagines a case of [Chase and Sanborn](#), both of whom were tasters for, and loved, Maxwell House

coffee.

After a while, both of them realize that they no longer love the same coffee.

Chase alleges that his qualia are the same but that his evaluations of those sensations has shifted.

Sanborn alleges that his qualia have changed.

If qualia are real, then there should be a fact of the matter about whether Chase and Sanborn are correct about their judgments.

But Dennett suspects that there is no fact of the matter, that it could be the case that Chase and Sanborn are just describing the same phenomena in different ways.

The challenge, for those who defend qualia, is to show that it is possible to determine whether they are correct or whether they are mis-diagnosing themselves.

Concerns about memory play a role in evaluating the Chase and Sanborn case.

Both tasters claim to be able to compare their former qualia with their present qualia.

We can perform tests on their memories, by examining their abilities to discriminate.

If they fail such tests, then we have reason to believe that their memories are faulty, and that their claims are unsupported.

But if they pass these tests, we still lack confirming evidence for their original claims.

We can also perform tests on their perceptual apparatus, their taste buds, to look for anomalies.

If Sanborn is right, there might be an obvious physiological explanation for the change in qualia, one which would not undermine their reality.

But, without a corresponding physiological explanation for the shift in qualia, there is no evidence to support Sanborn.

The general attack on the reality of qualia implicit in the Chase and Sanborn example relies on problems accessing our memories of qualia in order to perform intra-personal comparisons.

If qualia are inaccessible in memory, they will be useless to science, even if they are real.

If qualia are real phenomena for which physical science must account, we should be able to rely on them to discriminate between Chase and Sanborn.

But, Dennett argues, there are reasons to think that qualia are not substantial enough to play that role.

Dennett's argument mostly consists of characterizing the traditional notions of qualia and showing that there is nothing that has these characteristics.

He characterizes qualia as allegedly being ineffable (inexpressible); intrinsic; private; and directly or immediately apprehensible in consciousness.

Then, he argues that nothing really matches that description.

Consider drinking beer or wine.

Most everyone's first sip tastes awful.

But some people come to like it.

If tastes are acquired, then what we think about them shifts with our experience.

There is no single way that it tastes, independent of my past experiences.

In another example, imagine a pill which makes us like cauliflower.

Since we never liked cauliflower, it would seem that the pill must change its taste.

But another possibility is that it merely changes the way we feel about the taste.

The latter possibility is scientifically preferable, since it leaves the cauliflower alone.

But it also means, again, that there is no way that it tastes, independent of my past experiences.

Consider phenol-thiol-urea.

Some portion of humans find it extremely bitter and some find it tasteless.

Whether you find it bitter or tasteless depends on your genes.

If we got rid of all of the humans with a genetic ability to taste it, then we might start to believe that it is intrinsically tasteless.

If we did the reverse experiment, and got rid of all the humans with a genetic inability to taste it, then we might think that it is intrinsically bitter.

The taste is thus relational, rather than intrinsic.

As we have seen, the intrinsicity of secondary properties of objects has long been abandoned.

But, the question here is about the intrinsicity of the properties of my experience.

Dennett believes that we should abandon intrinsicity for qualia just as we abandoned it for public objects.

Dennett also cites worries about visual field's being intrinsically right-side up.

We can make spectacles which invert the field of vision.

It takes a few days, but subjects wearing inverting spectacles adapt.

Suppose we pressed them on this question: "Does your adaptation consist in your re-inverting your visual field, or in your turning the rest of your mind upside-down in a host of compensations?" If they demur, may we insist that there has to be a right answer, even if they cannot say with any confidence which it is?... Only a very naive view of visual perception could sustain the idea that one's visual field has a property of right-side-upness or upside-downness *independent of one's dispositions to react to it* - "intrinsic right-side-upness" we could call it (Dennett, "[Quining Qualia](#)" p 423).

There seems to be no fact of the matter about whether the subjects' visual fields are inverted and people adjust, or whether their field becomes resolved.

If Dennett's interpretation is correct, then the intrinsicity of my qualia seems lost.

To show that qualia are not really private, Dennett argues that third-person assessments of our qualia may be much better than first-person assessments.

In fact, Dennett takes this to be the moral of the memory problems that Chase and Sanborn faced.

In general, our own memories are liable to errors.

Videotape is much more reliable than memory, for most of us.

Dennett also points out that third-person assessments are better than first-person assessments in cases where we evaluate lighting intensity, or our own body temperature.

In such cases, we look to objective measures, over our own apprehensions.

If we had good evidence for the immediate, and infallible, apprehension of qualia, then we would override third-person considerations.

But, considerations of memory especially, seem to erode our confidence in our first-person access.

Empirical testing will not settle the Chase and Sanborn cases.

It just seems as if our qualitative states lack real, intrinsic properties.

We do not immediately hear all the harmonics, for a last example, in the vibration of a string.

But we can be trained to hear more of them.

Similarly, we can train ourselves to discriminate all sorts of tastes in wine.

Not only does our ability to train up our senses once again cast doubt on the intrinsic properties of our qualia, but it seems that these properties are neither directly or immediately apprehended.

We have subjective authority, in a limited sense, but not infallibility or incorrigibility.

III. The Subjective Character of Experience

So far, we have seen only the dualist's defense of the reality of first-person, introspective mental states. Two influential articles in the late twentieth century defended the irreducible reality of qualitative states. The first, in 1974, was by Thomas Nagel, and called "What Is it Like to Be a Bat?" Nagel asks us to imagine what it is like to be a bat. Bats echolocate using sonar; their perceptual apparatus is alien to us. We can do all sorts of physico-chemical studies on bats. Still, we will not be able to capture the subjective character of the bat's experience.

It will not help to try to imagine that one has webbing on one's arms, which enables one to fly around at dusk and dawn catching insects in one's mouth; that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals; and that one spends the day hanging upside down by one's feet in an attic. In so far as I can imagine this (which is not very far), it tells me only what it would be like for me to behave as a bat behaves. But that is not the question. I want to know what it is like for a bat to be a bat. Yet if I try to imagine this, I am restricted to the resources of my own mind, and those resources are inadequate to the task. I cannot perform it either by imagining additions to my present experience, or by imagining segments gradually subtracted from it, or by imagining some combination of additions, subtractions, and modifications. To the extent that I could look and behave like a wasp or a bat without changing my fundamental structure, my experiences would not be anything like the experiences of those animals... The best evidence would come from the experiences of bats, if we only knew what they were like (Nagel, "[What is it Like to Be a Bat](#)" p 439).

Similarly, super-intelligent [Martians can do all sorts of physico-chemical studies on us](#). But they won't be able to capture the subjective character of our experience.

If Nagel is right, there appear to be facts not expressible in human concepts. The third-person perspective, captured fairly by physicalist theories of mind, will be in principle insufficient for describing all the facts. There are facts that are not objective facts. Our conscious experience is essentially first-person. But, a physical theory is essentially third-person. Our experiences are ineffable. We can not describe experiences in a way that would make them available to others. Another way to put Nagel's problem is to say, contra Dennett, that qualia have intrinsic properties. Intrinsic properties are, by definition, not accessible from a third-person perspective.

IV. Epiphenomenalism

The second influential article defending qualia, in 1982, was Frank Jackson's "Epiphenomenal Qualia." Jackson's article has two parts. In the first part, Jackson opposes physicalism.

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 (Jackson 127).

Jackson claims that physicalism effectively omits two kinds of knowledge.

First, it omits the qualia.

Second, it omits the first-person perspective.

Nagel is right that the first-person perspective is missing.

But, every one's problem is no one's problem.

That is, there doesn't seem to be an objection to physicalism in the problem of imagining what it is like to be a bat, or any such alien creature.

The problem is just in the omission of the experience.

Jackson distinguishes his argument from Nagel's by denying that the real problem with physicalism has to do with the omission of the subjective point of view.

The problem, says Jackson, is not that physicalism omits the what-it-is-like.

The problem is that physicalism misses some facts.

If physicalism is wrong, something must replace it.

In the second part of his article, Jackson defends epiphenomenalism by considering three arguments for the falsity of physicalism based on the reality of qualia.

Epiphenomenalism is a dualist claim that allows causal efficacy only from the physical to the mental. On epiphenomenalism, physical events cause mental events, but mental events do not cause physical events.

Traditional epiphenomenalists may claim that mental states are completely non-efficacious.

Jackson's epiphenomenalism differs from the traditional version in two ways.

First, he remains agnostic whether mental states can affect, or cause, other mental states.

All epiphenomenalists hold that mental states cannot causally affect physical states.

My joy cannot cause me to jump up and down.

But perhaps it can cause me to have other joy-related mental states.

Second, Jackson's version of epiphenomenalism concerns mental properties, not mental states.

He argues that there are irreducibly mental properties of physical (brain) states, but not that there are non-corporeal minds.

That is, Jackson is defending a version of property dualism, not a version of substance dualism.

Consider physical theories of ordinary objects.

When we explain physical phenomena, like objects falling to the earth, we ignore the color of the object.

We reduce color to reflections of light.

But, light waves are not colors, sound waves are not sounds.

On the one hand, we have the phenomenal character of our experience.

On the other hand, we have objective physical properties.

With Locke, we say that sounds and colors (and all the conscious phenomenal characteristics) are just the secondary properties, their effects on us.

Only the primary qualities, the real qualities, matter to physical theory.

So, the primary/secondary distinction allows us to keep the phenomena, like color, by saying that there are terms in our physical theory for which our ordinary terms are shorthand.

In the case of qualia, though, recourse to the primary/secondary distinction will not help.

For, the sounds and colors, as qualia, are exactly what we need to explain about consciousness.

V. The Modal and Knowledge Arguments

Jackson insists that any theory of the mind must find a role for the qualities of our immediate experience. He calls himself a qualia freak.

He considers two distinct arguments for the legitimacy of qualia.

His modal argument is that zombies are possible.

My zombie is an organism just like me, except that it has no conscious experience.

There is a possible world with organisms exactly like us in every physical respect (and remember that includes functional states, physical history, *et al.*) but which differ from us profoundly in that they have no conscious mental life at all. But then what is it that we have and they lack? Not anything physical, *ex hypothesi*. In all physical regards we and they are exactly alike.

Consequently there is more to us than the purely physical (Jackson 130-1).

If zombies are possible, then it is possible for the same physical structure (or functional organization) to correspond both to a conscious person and to a zombie.

Thus, consciousness could not be explained by any physical properties of an organism (or functional structure).

Jackson points out that there is a disputed modal intuition at the core of the zombie question.

Some of us believe that zombies are possible; others don't.

Jackson says that no amount of physical information logically entails that another person is conscious.

Lots of ink has been spilled in recent years on the possibility of zombies.

See, for examples: [Zombies](#) [Zombies](#). Also: [Zombies](#).

If you don't believe that zombies are possible, then you might prefer the knowledge argument.

The knowledge argument has spawned a virtual [industry of discussion](#), especially of Mary.

Fred can discriminate between two different reds, red₁ and red₂, which look exactly the same to the rest of the normal-sighted world.

We can verify that he has this ability by looking at the physical, behavioral facts; he discriminates consistently between objects which reflect different wavelengths of light.

But we can not see the difference that he sees.

And all the physical facts won't tell us what that experience is, what the new colors are like.

It seems that physicalism leaves something out.

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 (Jackson 129).

The Fred case, like the zombie case, is science fiction.

One response to it is to question its plausibility.

Even if we can imagine such a case, we don't know if it is really possible.

We can, for example, imagine some aspects of a world without gravity.

But such a world is not possible, at least in some sense.

Part of the importance of such cases, for philosophers, has been to push people to think more clearly about possibility and its connection to conceivability.

Jackson's Mary case has been more widely discussed.

Mary knows all the physical facts about color, while living in a completely black-and-white world.

She knows all about wavelengths of light, the reflective and absorbing properties of objects, details about retinas and rods and cones, and all the facts about neural processing of visual images.

She knows that real bananas are yellow and that the real sky is blue, but she only sees black-and-white versions of these objects.

When she leaves her room, she sees colors for the first time.

On the one hand, she knows, prior to going into the world, all there is to know about color.

On the other hand, she seems to learn something when she leaves.

Jackson concludes that we can have all the scientific knowledge that there is to have, and still learn something about qualia.

VI. Qualia and Causal Efficacy

If qualia are real, then they appear to be causally efficacious.

My sense experiences affect both my actions and my other mental states.

Thus, Jackson's argument seems to lead to Cartesian-style dualism which Jackson wishes to avoid.

Intuitions about the causal efficacy of mental states constituted a central criticism of behaviorism, and a central motivation for both identity theory and functionalism.

If mental states are just physical states, as any token physicalist has it, then causal efficacy is welcome.

But, if mental properties are non-physical, their causal efficacy is suspicious.

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 (Jackson 128).

Jackson thus must find room for a position on which qualia are real but not causally efficacious.

He develops this position in response to arguments for the dualist's claim that the reality of qualia entail their causal efficacy.

It seems, to the substance dualist, that the pain I feel when a piano drops on my foot causes me to hop about and bark.

Jackson argues that qualia are causally inert byproducts of causal interactions.

All causation appears at the physical level.

The epiphenomenalist's mental properties just come along for the ride.

Jackson provides three arguments against the causal efficacy of epiphenomenal mental properties: the Hume Argument; the Darwin Argument; and the Other Minds Argument.

Hume argues that we posit causal connections on the basis of our experiences of conjunctions of events.

For example, every time I let go of my keys, they fall to the ground.

We posit some underlying cause of the phenomenon of the constant conjunction: gravity.

Consider the fact that every time my watch says the time, my cell phone says that it is the same time.

We have a reliable conjunction of events.

But we do not posit a causal relation between my watch and my cell phone.

Rather, we find common underlying causes.

It is not that my watch makes my cell phone say that it is a particular time, or vice versa.

It is just that the laws of physics work reliably in both cases.

Applying this lesson, if there were some underlying cause of both my pain and my hopping about, we could eliminate the belief that qualia were causally efficacious.

Jackson's movie example is supposed to show that we can avoid admitting the causal efficacy of qualia on the basis of examples like that of the piano.

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 behavior. It is simply a consequence of the fact that certain happenings in the brain cause both (Jackson 133).

It appears that there is a causal link between the image of Lee Marvin's fist and the image of John Wayne's head.

The causal link is not properly located in the images, but in some underlying causes of these images.

To consider the Darwin argument against causal efficacy, consider an evolutionary argument for the causal efficacy of qualia:

- DA DA1. We have qualia.
- DA2. Lower animals, earlier forms of life, do not.
- DA3. So, qualia appeared at some point in evolution.
- DA4. Thus, qualia must have some evolutionary role.
- DAC. Thus, qualia must be causally efficacious.

Jackson points out that evolutionary explanations such as this are invalid, since some traits which are not conducive to survival may persist, as long as they do not hinder survival too much.

He uses the example of the heavy coat of a polar bear.

Thus, all we can conclude from DA1 - DA4 is that qualia are either causally efficacious, or byproducts of something that is causally efficacious.

Just as in the Hume argument, causal efficacy need not be ascribed directly to the qualia.

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 is irrelevant to survival (Jackson 135).

Lastly, the argument from other minds that we should attribute causal efficacy to qualia runs as follows:

- OM OM1. We infer know that other people have minds because, at least in part, of their behavior.
- OM2. But, in my own case, it seems that my qualia cause my behavior.
- OM3. It is reasonable to posit that the same causal relation holds in the cases of other people.
- OMC. So, we should attribute causal efficacy to the qualia.

Again, Jackson argues that the causal efficacy should be attributed to whatever causes the qualia, and not to the qualia themselves.

We can attribute qualia to other people just because they have behavior which correlates with qualia in

my case.

But, we need not think that the qualia cause anything, even in our own case.

Qualia may be caused by something casually efficacious, but they need not be themselves involved.

VII. Two Paths

We are at a fork in the road.

One path accepts the reality of qualia, but denies the sufficiency of physicalism.

The other path takes physicalism as brute, but omits aspects of conscious experience.

On the qualia path, we accommodate phenomenal facts.

Then, the insufficiency of physicalism seems to infect all kinds of reductive definitions.

For example, we reduce color to wavelength of reflected light.

But descriptions of the wavelength of light omit facts about its effects on perceivers.

If we are looking for a scientific theory, which includes explanations of consciousness and conscious experience, we can not dismiss these facts.

So, the qualia path may not only force us toward some kind of dualism.

It may also undermine all kinds of physical reductions.

On the physicalist path, we have gotten stuck looking for appropriate type-identities for mental states. Type-identity statements (what it is to be a thing of a certain type) must be made with reference to the appropriate regularities.

Type-identity of elementary particles will be made in terms of charge, because, presumably, charge is an element of the basic physical laws.

Type-identity of species will be made in terms of genetic constitution.

Type-identity of water will be made in terms of molecular constitution.

Similarly, it seems that type-identity of mental states must be made in terms of psychological laws.

But psychological laws are most naturally formulated in terms of inner states which appear to be difficult for the physicalist to capture.

That is the essence of the criticism of the identity theory for lacking a relational account of mental states.

We have been considering attempts to define the mind by looking for acceptable type-identity statements in which one side corresponds to our ordinary psychological states, to the terms of folk psychology.

The Cartesian sorts mental states in the right way, according to psychological regularities which hold among our mental states.

But, since we lack third-person access to the dualist's mental states, the Cartesian lacks key elements of a scientifically legitimate theory: verifiability, replicability, etc.

The behaviorist seeks type-identities in terms of observable behavior, but they have problems sorting mental states, since they do so according to observable criteria, which do not do justice to the internal states.

The identity theorist sorts mental states in terms of brain states, which leads to chauvinism.

Functionalists seem to have an advantage over these other positions, because they sort mental states according to their causal roles, in terms of functional states of a probabilistic automaton.

But functionalism appears too liberal.

Given the successive failures of each program, perhaps we should give up trying to reduce mental states at all.