

Mind and Illusion

FRANK JACKSON

Much of the contemporary debate in the philosophy of mind is concerned with the clash between certain strongly held intuitions and what science tells us about the mind and its relation to the world. What science tells us about the mind points strongly towards some version or other of physicalism. The intuitions, in one way or another, suggest that there is something seriously incomplete about any purely physical story about the mind.

For our purposes here, we can be vague about the detail and think broadly of physicalism as the view that the mind is a purely physical part of a purely physical world. Exactly how to delineate the physical will not be crucial: anything of a kind that plays a central role in physics, chemistry, biology, neuroscience and the like, along with the *a priori* associated functional and relational properties count as far as we are concerned.

Most contemporary philosophers given a choice between going with science and going with intuitions, go with science. Although I once dissented from the majority, I have capitulated and now see the interesting issue as being where the arguments from the intuitions against physicalism—the arguments that seem so compelling—go wrong.¹

For some time, I have thought that the case for physicalism is sufficiently strong that we can be confident that the arguments from the intuitions go wrong somewhere, but where is somewhere?

This paper offers an answer to that question for the knowledge argument against

physicalism. I start with a reminder about the argument. I then consider one very popular way of dismissing it and explain why I am unmoved by it. The discussion of this way delivers a constraint that any satisfying physicalist reply to the knowledge argument should meet. The rest of the paper gives the answer I favour to where the knowledge argument goes wrong. This answer rests on a representationalist account of sensory experience and, as the title of the paper indicates, I say, among other things, that there is a pervasive illusion that conspires to lead us astray when we think about what it is like to have a colour experience.

The knowledge argument²

The epistemic intuition that founds the knowledge argument is that you cannot deduce from purely physical information about us and our world, all there is to know about the nature of our world because you cannot deduce how things look to us, especially in regard to colour. More general versions of the argument make the same claim for all the mental states with a phenomenology—the states for which there is something it is like to be in them—as it is so often put, and sometimes for consciousness. But we will be almost entirely concerned with colour experiences. We will say nothing about consciousness *per se*; our concern is with the phenomenology of visual experience, and not our consciousness of it or of mental states in general.

The familiar story about Mary is a way to make vivid and appealing the claim about lack of deducibility. To rehearse it ever so briefly: A brilliant scientist, Mary, is confined in a black and white room without windows. She herself is painted white all over and dressed in black. All her information about the world and its workings comes

from black and white sources like books without coloured pictures and black and white television. She is, despite these artificial restrictions, extraordinarily knowledgeable about the physical nature of our world, including the neurophysiology of human beings and sentient creatures in general, and how their neurophysiology underpins their interactions with their surroundings. Can she in principle deduce from all this physical information, what it is like to see, say, red?

There is a strong intuition that she cannot. This intuition is reinforced by reflecting on what would happen should she be released from her room. Assuming that there is nothing wrong with her colour vision despite its lack of exercise during her incarceration, she would learn what it is like to see red, and it is plausible that this would be learning something about the nature of our world, including especially the nature of the colour experiences subjects enjoy. From this it would follow that she did not know beforehand all there was to know about our world.

Moreover, there is a marked contrast with our epistemic relation to properties like solidity, elasticity, boiling, valency and the like. If I give you enough information about the behaviour of a substance's molecules and how they govern the substance's interactions with its environment, you will be able to work out whether it is a liquid, a solid or a gas. If I tell you about the forces that hold water molecules together and the way increases in the velocity of those molecules as a result of heating can lead to these molecules reaching escape velocity, you will learn about boiling.³ Likewise for valency and elasticity. But the deduction of what it is like to see red from purely physical information seems a totally different matter.

There are two challenges to physicalism here. One is to explain why there should be a marked apparent difference between the case of seeing red and the case of being liquid or boiling. After all, the phenomena are *alike* in being purely physical ones according to physicalism. The second, more direct challenge is to explain how it can be that Mary's knowledge of our world's nature is, it seems, deficient, despite the fact that she knows all there is to know according to physicalism.

I now turn to the very popular response to the knowledge argument that seems to me to fail but which gives us a constraint on any acceptable physicalist response.

The response that draws on *a posteriori* necessity

This response⁴ on behalf of physicalism to the knowledge argument starts from the point that being necessitated does not imply being *a priori* derivable. This suggests that although physicalists are committed to the experiential being necessitated by a rich enough physical account of our world—otherwise it would take more than the physical nature of our world to secure its experiential nature, contrary to physicalism—they are not committed to the experiential being *a priori* derivable from the physical. But the epistemic intuition that lies behind the knowledge argument is, when all is said and done, that Mary cannot carry out an *a priori* derivation from the physical information imagined to be at her disposal to the phenomenology of colour vision. Physicalists should respond to the knowledge argument by adopting a version of physicalism according to which the experiential is necessitated by the physical but is not *a priori* derivable from the physical.

I have two reasons for rejecting this reply. The first I have given a number of times. It draws on the two-dimensional account of the necessary *a posteriori*. I will not repeat it here.⁵ My second reason can be introduced by reflecting on the famous reduction of the thermodynamic theory of gases to the kinetic theory via statistical mechanics.

Our belief that gases have temperature and pressure is grounded in their behaviour. Moreover, we know that their behaviour is fully explained by the various features recognised and named in the kinetic theory of gases. There is no need to postulate any extra features of gases in order to explain their behaviour. This makes it very hard to hold that no matter how much information we have framed in the terms of the kinetic theory and in terms of the functional roles played by the properties picked out by the terms of that theory, and no matter how confident we are that the kinetic theory and its future developments provide a complete picture in the relevant respects of the essential nature of gases, the passage from this information to whether or not gases are hot and have pressure is *a posteriori*. What relevant information are we waiting on? We know that all we will get is more of the same. Scepticism about gases having temperature and pressure threatens if we insist that we cannot go *a priori* from the molecular account of gases and the concomitant functional roles to gases having temperature and pressure.

This point is implicit in the well-known schematic account of why it is right to identify temperature in gases with mean molecular kinetic energy:

Temperature in gases is that which does so and so (*a priori* premise about the concept of temperature)

That which does so and so is mean molecular kinetic energy (empirical premise)

Therefore, temperature in gases is mean molecular kinetic energy.

The need for the first, *a priori* premise is sometimes challenged.⁶ But unless something like the first premise is *a priori*, eliminativism about temperature and pressure in gases is inevitable. The right conclusion from the discoveries of the kinetic theory of gases could only be that gases are not hot on the ground that what we needed temperature to explain (their feeling hot and behaving thus and so) is fully explained by their mean molecular kinetic energy. *Mutatis mutandis* for pressure.

It is sometimes objected to this argument that identities need no explanation. I doubt this doctrine.⁷ But the issue is by the way. Identities certainly need justification, and the problem is that we have a choice between

- (A) Temperature is not a property of gases although there is plenty of molecular kinetic energy, and the mean value of that does all the explaining of gas behaviour once assigned to having such and such a temperature.
- (B) Having such and such a temperature in gases is one and the same property as having so and so a mean molecular kinetic energy, and 'they' do all the needed explaining of gas behaviour.

Without the first, *a priori* premise above, we have no reason to favour (B) over (A).

The considerations that tell us that we had better be able to move *a priori* from the molecular account of gases to the temperature account can be generalised to the

question of all of our empirical beliefs about what our world is like. Physicalists hold (have to hold) that the evidence we have for any of our claims about what our world is like—that England fought two World Wars, that horses eat grass, that Carter was a one-term President of the United States, and so on—is determined without remainder by our world's physical nature. How then can we be justified in holding that we have evidence for what our world is like that outruns what might be inferred in principle from its physical nature alone? It might be objected that this rhetorical question assumes an unduly 'causal cum best explanation' view of the relation of evidence to empirical hypothesis. What about simplicity and all that? But physicalists hold that considerations of simplicity, good methodology, and all the rest, favour physicalism. That is why they are physicalists (and rightly so, in my view).

It is this wider consideration that explains my puzzlement over why many hold that the claim that physicalism is committed to the *a priori* deducibility of the way the world is in all empirical respects from the physical nature of the world is an extreme one.⁸ Think of the famous 'Russell hypothesis'. According to it, the world came into existence five minutes ago containing each and every putative 'trace' that might suggest that it has existed since the big bang. As a result, we cannot here and now point to features that distinguish the correct view that our world has existed since the big bang from the Russell hypothesis. What entitles us to reject the Russell hypothesis is that it violates the principles of good theory construction by being *excessively ad hoc*. Now consider the *bare physicalism hypothesis*: the hypothesis that the world is exactly as is required to make the physical account of it true in each and every detail but nothing more is true of this world in the sense that nothing that fails to follow *a priori* from the physical

account is true of it. This hypothesis is not *ad hoc* and has all the explanatory power and simplicity we can reasonably demand. Ergo, we physicalists can have no reason to go beyond the bare physicalist picture.

It might (will) be objected that bare physicalism is *a posteriori* impossible, that there are empirical truths about our world, including truths about experiences, that are necessitated by the bare physical account but which do not follow *a priori* from that account. But that would be to miss the point. The point is that we could not know this. Bare physicalism is a conceptual possibility; the argument is that we have no reason not to allow that it is also a metaphysical possibility. Or, to make the point with a word, recall that many call conceptual possibility *epistemic* possibility. Or, to make the point with an example, those who hold that the existence of God is *a posteriori* necessary (or *a posteriori* necessitated by agreed features of our world) are not thereby excused from having to provide reasons for believing in God.

To avoid misunderstanding, I should emphasize that when I talk of being able to move *a priori* from the physical account to, say, Carter being a one-term President, I do not mean being able to move literally. I mean that there exists an *a priori* entailment. We cannot derive the gravitational centre of the universe from the mass and location of all its parts because, first, we do not and could not know the mass and location of all its parts, and, secondly, the calculation would be way beyond our powers. All the same, the location of the centre of gravity does follow *a priori* from the physical account of our world, and we know that it does.

This gives us the following constraint on any physicalist solution to the challenge of the

epistemic intuition: it should allow us to see how the passage from the physical to the nature of colour experience might possibly be, somehow or other, *a priori*.

I now come to the positive part of the paper; the part where I explain why physicalists are entitled to reject the epistemic intuition. As heralded, my argument will involve the claim that we are under an illusion about the nature of colour experience, an illusion that fuels the epistemic intuition.⁹

Mistaking intensional properties for instantiated properties

I start with the diaphanousness of experience: G. E. Moore's thesis that the qualitative character of experience is the character of the putative object of experience.¹⁰ The redness of sensings of red is the putative redness of what is seen; when vision is blurred, what is seen appears to be blurred; the location quality of a sound is the putative location of the sound; the experience of movement is the experience of something putatively moving; and so on. Hume observes that the self's experiences always get in the way of experiencing the self.¹¹ Equally, the putative properties of what is experienced always get in the way of accessing the qualities of experience. I am going to take diaphanousness for granted. The case for it is widely accepted¹² and it is especially appealing in the case of our topic, colour experience. Indeed, reservations about it are typically confined to certain bodily sensations where attitudes, pro or con, arguably contribute to the felt quality. The degree to which we dislike a pain is arguably part of its feel.

There are two very different ways to think of the lesson of diaphanousness,

corresponding to two very different ways of thinking of the object that putatively has the qualities. On one, Moore's, the object really is an object. It is the object of the act-object theory of experience or the sense datum theory of sensing: experiences are composed of an act of awareness directed to an object or sense datum which bears the qualities. And the lesson of diaphanousness is that these qualities determine the qualitative nature of the experience. On the other way of thinking, Harman's, *op. cit.*, for example, the object is an intensional object. That is to say, 'it' is not an object at all, and our use of verbal constructions that belong in the syntactic category of names is a convenient, if metaphysically misleading, way of talking about how things are being represented to be. We talk of being directly aware of a square shape in our visual fields but there is no square shape to which we stand in the relation of direct awareness; rather, our visual experience represents that there is something square before us. What makes it right to use the word 'square' in describing our experience is not a relation to something which has the property the word stands for but the fact that the way the experience represents things as being can only be correct if there is something square in existence. The squareness of an experience is an intensional property, not an instantiated one. The same goes *mutatis mutandis* for all the properties we ascribe to what is presented in experience, the properties we have in mind when we talk of the qualities of experience and to which the argument from diaphanousness applies.¹³ When we use words like 'square', 'two feet away' and 'red' to characterise our experiences, we pick out intensional properties not instantiated ones.

I think, with the current majority, that the second is the right way to think of the lesson of diaphanousness. My reason is that perceptual experience *represents*. My experience

as of a round, red object in front of me represents that there is a round, red object in front of me. I may or may not accept that things are as they are being represented to be, but I take it as axiomatic that each and every sensory experience represents that things are thus and so.

This implies that the first way of thinking of the lesson of the argument from diaphanousness, the way that leads to the sense datum theory, must be rejected. We one-time sense datum theorists thought that the requirement that there be something red and round, say, of which the subject is directly aware, automatically captures, or part way captures, the key representational notion.¹⁴ This is a mistake. It is true that I can represent how I am representing something to be by using the actual way something is. For example, I might represent to you the colour I remember *X* to be by holding up an actual sample of the colour. Here I would be using the actual colour of one thing, the sample, to represent how my memory represents the colour of something else to be; a colour which *X* may or may not have. In that sense, we have a model for understanding the sense datum theory. But, and this is the crucial point, the fact that I am using an actual sample of the colour cuts no representational ice *per se*. I could be using the sample to indicate the one colour I do *not* think *X* has. Or I could be following the convention of holding up a sample with the colour *complementary* to that I remember *X* as having. In the same way, standing in a certain direct-awareness relationship to a mental item with such and such properties says nothing, represents nothing *per se*, about how the world is. The act-object cum sense datum theory leaves out the most important feature of experience: its essentially representational nature. In order to capture the representational nature of perception, what makes it true that words like 'red' and

'square' apply to our experiences has to be understood on the intensional model.

It might be objected that this argument from the fact that perception represents leaves open the possibility that some but not all of the properties of experience are intensional.

Why not hold that experiences have both a representational aspect and a non-representational aspect?¹⁵ In a sense they do. It may be a fact about an experience that it is occurring in Alaska or in the Middle Ages, and neither of these properties is an intensional property of the experience. But the issue for us is whether the aspects that constitute the phenomenal nature of an experience outrun its representational nature, and there are good reasons to deny this.

First, whenever there is a difference in phenomenal character, there is a difference in how things are being represented to be. This follows from diaphanousness. Any change in phenomenal character means a change in the putative character of what is being experienced, and a change in the putative character of what is being experienced *is* a change in how things are being represented to be. Make an experience of red a bit brighter and you make it the case that your experience represents that some object's redness is that bit brighter. But if phenomenal character outran representational character, it would be possible to change the former and leave the latter unchanged.

There have, of course, been attempts to describe cases where phenomenal character differs without a difference in representational content and an important exercise is the critical review of all the cases that might be thought to show the possibility of phenomenal variation without difference in representational content. I am not going to conduct this review, because I think the job has been well done by other supporters of

representationalism.¹⁶

Secondly, there is a marked contrast between, on the one hand, the way representational devices like maps and sentences represent, and, on the other, the way perceptual experience represents. There is a gap between vehicle of representation and what is represented in the first case that does not exist in the second. In the case of maps and sentences, we can distinguish the features that do the representing—the gap between the isobars on a weather map, the concatenation of the letters ‘c’, ‘a’ and ‘t’ in that order in a sentence, the green colouring on parts of a map, *etc.*—from what they represent: a pressure gradient, a cat, areas of high rainfall, *etc.* We can, for example, describe the gap between the isobars without any reference to what it represents. But, in the case of perceptual experience, we cannot. When I have a visual experience of a roundish, red object in front of me, *that* is what it represents. My very description of the vehicle of representation delivers how it represents things to be. I may or may not accept that things are the way they are being represented to be, but there is just the one way that things are being represented to be, and that way is part and parcel of the quality of the experience. Ergo, we have to understand the qualities of experience in terms of intensional properties.¹⁷

A major issue for the intensionalist account is how to distinguish sensory representational states from more purely cognitive representational ones like belief. But rather than break the flow of the argument, I postpone my discussion of it. In the next few sections we take the intensionalist picture as a given and note how it allows physicalists to explain away the epistemic intuition.

Explaining away the epistemic intuition

We start by noting how the intensional account undermines the picture of experience that goes with the phrase 'what it is like'.

There is a redness about sensing red (a yellowness about sensing yellow, and so on). We naturally think of the redness as a property we are acquainted with when we sense red and as the property Mary finds out about on her release. We may want to distinguish redness as a property of objects from redness as a property of an area of our visual field, perhaps using 'red*' for the latter. Either way, what it is like is, on the picture, a matter of having redness or redness*, knowing what it is like is knowing about redness or redness*, and the knowledge argument is an argument to the conclusion that Mary does not know about redness or redness*—that is, about the property we are, according to the picture, acquainted with when we sense red.

Intensionalism tells us that there is no such property. To suppose otherwise is to mistake an intensional property for an instantiated one. Of course, when I sense red and you sense red, there is something in common between us which we English speakers report with descriptions that include the word 'red'. But what is in common is not the property tagged with the word 'red' but, first, how things have to be for our experiences to represent correctly, and, second, our both being in states that represent things as being that way.

Intensionalism means that no amount of tub-thumping assertion by dualists (including by me in the past) that the *redness* of seeing red cannot be accommodated in the austere

physicalist picture carries any weight. That striking feature is a feature of how things are being represented to be, and if, as claimed by the tub thumpers, it is transparently a feature that has no place in the physicalist picture, what follows is that physicalists should deny that anything has that striking feature. And this is no argument against physicalism. Physicalists can allow that people are sometimes in states that represent that things have a non-physical property. Examples are people who believe that there are fairies. What physicalists must deny is that such properties are instantiated.

Moreover, the representationalist-cum-intensionalist approach can explain the origin of the dualist conviction that redness is non-physical. It is vital for our survival that we are able to pick out recurring patterns. Recognising your best friend or a hungry tiger requires spotting a commonality. Sometimes these patterns are salient ones. Square tables have an obvious commonality. Sometimes they are not. An example is the commonality that unites an acceptable pronunciation of a given word in English. The lack of salience is why it is hard to develop speech to text computer programs, though the fact that it is nevertheless possible, that we always knew it was possible in principle and now know that it is possible in practice, reminds us that it is a folk view that there is a commonality. In many cases, the commonality's importance lies in highly relational facts about it. If the theory that colour vision evolved as an aid to the detection of food is correct, a series of highly unobvious optical commonalities between edible things and differences from their forest backgrounds are the patterns colour vision evolved to detect. Now, highly unobvious commonalities like these normally get detected only after a great deal of collecting and bringing together of information. Colour experience is, therefore, a quite unusually 'quick' way of acquiring highly unobvious relational and

functional information. It is, in this regard, like the way we acquire information about intrinsic properties: one look at an object tells us that it is more or less round. In consequence, colour experience presents to us as if it were the acquisition of information about highly salient, more or less intrinsic features of our surroundings. But there are no physical features fitting this characterization; in consequence, colour experience presents itself to us as if it were information about certain non-physical features. Indeed, we may want to go so far as to say that sensing red *misrepresents* how things are. If this is right, we should say that nothing is red, for nothing would be as our experience of red represents things as being; we should be eliminativists about red and about colour in general. A more moderate position is that though our experience of colour contains a substantial degree of misrepresentation—the misrepresentation that leads dualists astray—there are complex physical properties ‘out there’ which stand in relations near enough to those captured by the colour solid for us to be able to identify them with the various colours.

Meeting the constraint

We argued that the physicalist response to the epistemic intuition should allow us to see how the nature of colour experience might possibly follow *a priori* from the physical account of what our world is like. The representationalist account of sensory experience meets this constraint.

Seeing red is being in a certain kind of representational state on this account. The project of finding an analysis of representation is not an easy one—to put it mildly. But it is a standard item on the philosophical agenda and the answers that have been, or are

likely to be, canvassed are all answers that would allow the fact of representation to follow *a priori* from the physical account of what our world is like. They are accounts that talk of co-variation, causal connections of various kinds, selectional histories, and the like, and accounts made from these kinds of ingredients are ones that might be determined *a priori* by the purely physical.

We will need also an account of how sensing red represents things as being—the content. This would involve, *inter alia*, making up our mind on whether or not the content of sensing red was such as to imply, given physicalism, eliminativism about redness. Again, this will not be easy—if it were, it would have been done long ago—but there is no reason to think it would be an account that would make being in a state that has that representational content something that could not be derived *a priori* from the physicalist picture of what our world is like. I know only too well the residual feeling that somehow the *redness* could not be got out of the physical picture alone, but that is nothing more than a hangover from the conflation of instantiated property with intensional property. That 'redness' is not a feature one is acquainted with, but instead is a matter of how things are being represented to be.

What happens to Mary on her release?

The epistemic intuition is that it is impossible to deduce what it is like to sense red from the physical account of our world. In particular, Mary in her room will not be able to do it. I have argued that if what it is like means all the properties of seeing red, it is possible in principle to deduce them all. That follows from representationalism, and the appearance to the contrary arises from the conflation of intensional properties with

instantiated ones. But this leaves open what to say of a positive kind about what would happen to Mary on her release. The negative points that she would not learn about a feature of our world she could not know of while incarcerated, and that tub thumping convictions to the contrary carry no weight, do not tell us the positive side of the story.

What to say about the relevant change in Mary turns, it seems to me, on what to say about an old, hard issue for representationalist approaches to experience. It is the issue we postponed earlier of how to find the feel in the representationalist picture.

Sensing and believing

I can believe that there is, here and now, a round, red object in front of me without having the relevant visual experience. Perhaps my eyes are shut but I remember, or perhaps I am being told, that there is such an object here and now in front of me. Or perhaps the thought that there is such an object in front of me has simply ‘come into my mind’, and I have boldly gone along with it. Or perhaps I am one of the blind sighted: it seems like guessing but my success rate shows that I am drawing on a subliminal representational state.

It can be very tempting at this point to try for a mixed theory. Sensory experiences have a representational component and a sensory one. The difference between belief *per se* and sensory experience lies in a sensory addition. But we saw the problems for this earlier. For example, if this is the right view to take, it should be possible to vary the sensory part alone, but for every sensory difference, there is a representational difference. Moreover, it is hard to make sense of a non-representational, sensory core.

Any experience with some 'colour or shape feel' is putatively of something coloured and shaped somewhere, and thereby represents something about that location.¹⁸ Once there is some phenomenal experiential nature, there is thereby some representation.

Conceptual versus nonconceptual content—a wrong turn

Many representationalists tackle the problem of finding the 'feel' via a distinction between conceptual and nonconceptual content. The claim is that belief has conceptual content, whereas experience has nonconceptual content.¹⁹ I think that there are problems for this style of response.

The view that beliefs have conceptual content whereas experiences have nonconceptual content can be understood in two different ways.²⁰ It can be thought of as the view that beliefs and experiences have content in different senses of 'content'; that they have different kinds of content in the strong sense in which *that there are electrons* and *that there are protons* do *not* automatically count as different kinds of content. I think this is the wrong way for representationalists to go.²¹ Belief is the representational state *par excellence*. This means that to hold that experience has content in some sense in which belief does not is to deny rather than affirm representationalism about experience. There needs to be an univocal sense of 'content' at work when we discuss representationalism; a sense on which content is how things are being represented to be, and on which both beliefs and experiences have (representational) content.

Of course, many say that the content of belief (and thought more generally) is a structured entity containing concepts.²² But this should not, it seems to me, be

interpreted so as to run counter to what we have just said. When I believe that things are as my experience represents them to be, what I believe is precisely that things are as my experience says that they are, not something else.²³ Alex Byrne says ‘that the content of perception ... may outrun the representational capacity of thought ... is surely the default assumption’.²⁴ But we can *think* that things are *exactly* as our experience represents them to be. What is outrun is our capacity to capture the content in words, but that is another question. As it happens, my current experience correctly represents that there is something rectangular before me. I also believe that there is. What makes my experience correct and my belief true is the very same configuration of matter in front of me, and that configuration contains no concepts. Maybe, in addition, my belief implies that I stand in some special relation to concepts, but that would be a reason to acknowledge an additional content to the representational one that belief and experience both possess. And you might, or might not, hold that sensory states possess that kind of content also. But giving belief a possibly extra kind of content is not going to help us with our problem. Our problem is that both belief and experience represent that things are thus and so but only experience has feel—or anyway feel to the relevant degree. We are looking for something extra, so to speak, for experience, not a possible extra for belief.

The second way of understanding the view that belief has conceptual content whereas experience has non-conceptual content is as a claim about what it takes to have a belief with (representational) content versus what it takes to have an experience with (representational) content.²⁵ The kinds of content are the same but what it takes for the states to have them differs. Experience represents in a way that is independent, or

largely independent, of subjects' mastery of concepts, whereas belief does not. For example, it is observed that we can perceptually discriminate many more colours than we have names for or can remember. It is then inferred that I might have a perceptual state that represents that something is, say, red_{17} , without having the concept of red_{17} . But I could not believe that something is red_{17} without having the concept of red_{17} .

I doubt the claim that perceptual representation is nonconceptual in the explained sense. To perceptually represent that things are thus and so essentially involves discrimination and categorisation, and that is to place things under concepts.²⁶ Of course, I agree that when I experience red_{17} , I need not have the term ' red_{17} ' in my linguistic repertoire; I need not be representing that the colour before me is correctly tagged ' red_{17} '; it need not be the case that before I had the experience, I had the concept of red_{17} ; and my ability to remember and identify the precise shade may be very short-lived. But none of these points imply that I do not have the concept of red_{17} at the time I experience it.²⁷ When I learn the right term for the shade I can see, namely, the term ' red_{17} ', it will be very different from learning about momentum, charm in physics or inertial frames, which do involve acquiring new concepts. It will simply be acquiring a term for something I already grasp. My tagging the shade with the word does not create the concept in me though it does give me the wherewithal to *say* that it applies.²⁸ Any thought to the contrary would appear to conflate the concept of red_{17} —the shade—with the distinct, relational concept of being indistinguishable from the sample labelled ' red_{17} ' in some colour chart. It might be objected that this latter concept is the one we have in fact been talking about all along. But if this is the case, the initial datum that we experience red_{17} prior to acquaintance with colour charts is false. Prior to acquaintance with colour

charts, we do not experience colours as being the same as such and such a colour on a colour chart.

The same goes for shapes. It is sometimes suggested that when presented with a highly idiosyncratic shape, you may experience it but not have the concept of it. But we need to distinguish two cases. In one, you see something as having the highly idiosyncratic shape but lack a word for it. In this case, you do have the concept. All that is lacking is a word for it, which you can remedy by making one up for yourself or by asking around to find out if there is already one in, say, English.²⁹ In the second kind of case, you do *not* experience the shape prior to having the word and the concept. There are cases where you see that something has some complex shape or other, where that shape is in fact *S*, but fail to see it *as S*. You are then told the right word for the shape, acquire the concept it falls under, and thereby acquire the ability to see it *as S*. But then it is false that your experience represented that something is *S* prior to your mastery of the concept. Your acquisition of the concept changes the perceptual experience.

Of course, what it is to have a concept is disputed territory and one might define concept possession in terms of having a word for that which falls under the concept. But in that case many beliefs lack conceptual content—animals and people have beliefs for which they do not have words. Or, more generally, one might raise the bar on what it is to possess a given concept in a way that, although it is plausible that anyone who believes that something is *K* has the concept of *K*, it is not plausible that anyone whose experience represents that something is *K* has the concept of *K*. But it is hard to see how any such reading of what it is to possess a concept could help with our problem. It *adds* to what it takes to believe, and, as we noted earlier, we are looking for something

experience has and belief *per se* lacks.

A different way of finding the feel

To find the 'feel', I think representationalists should ask what is special about the representation that takes place when something looks or feels a certain way.³⁰ It seems to me that there are five distinctive features of cases where our sensory experience represents that things are thus and so.

First, such representation is rich. Visual experience represents how things are here and now in terms of colour, shape, location, extension, orientation and motion. Tactual experience represents how things are in terms of shape, motion, texture, extension, orientation and temperature.

Secondly, it is *inextricably* rich. A sentence that says *X* is red and round, has a part more concerned with redness and a part more concerned with roundness, and we can use sentences to represent something about colour while being completely silent about shape or motion or position, and conversely. But you cannot prise the colour bit from the shape bit of a visual experience. In representing something about shape, a visual experience *ipso facto* says something about colour (in the wide sense that includes white, black and grey); and a similar point applies to extension, location and motion. Equally, you cannot prise the texture and temperature bits from the shape bit of a tactile representation. Something cannot feel to have some shape or other without feeling to have a texture or a temperature (in the wide sense that includes being neither hotter nor colder than one's limb).

Thirdly, the representation is immediate. Reading from a piece of paper that there is something of such and such a colour, location, *etc.* typically induces a belief that represents that there is, but it does so *via* representing that there is a piece of paper with certain marks on it. Of course, immediacy may vary over time. Someone who uses a stick to feel the shape of an object down a hole will start by working from the feel of the end of the stick in their hand but typically ends up over time in a state that represents immediately the object's shape. The transition will match the transition from having an experience they characterise in terms of how their hand is felt to be to one they characterise in terms of how the object at the end of the stick is being felt to be.

Fourthly, there is a causal element in the content. Perception represents the world as interacting with us. When I hear a sound as being, say, behind and to the left, my experience represents the sound as *coming from* this location. To feel something is to feel in part its *contact with* one's body. Vision represents things as being located where we *see* them as being, as being at the location from which they are affecting us via our sense of sight.

Finally, sensory experience plays a distinctive functional role. Many years ago Armstrong analysed perceptual experience in terms of the acquisition of a disposition to believe as a result of the operation of one's senses.³¹ But, as many have objected, the top line in the Müller-Lyer figure looks longer despite the fact that, for experienced customers, there is no tendency whatever to believe that it is. What is, however, true of sensory experience is that it plays a distinctive functional role in mediating between one state of belief and another. It is not itself a state of belief. And it need not move a subject into a state of belief that represents as it does—the subject may know that they

are the subject of illusion or hallucination, or may already believe things are as the experience represents them—but it will determine a function that maps states of belief onto states of belief. A subject's posterior state of belief supervenes on their prior state of belief conjoined with their sensory experience.

Obviously, there is much more to say here, both by way of elucidation and by way of defence, but I hope the leading idea is clear. It is that if a representational state's content has inextricably and immediately the requisite richness, and if the state plays the right functional role, we get the phenomenology for free. In such cases, there must be the kind of experience that the blind sighted, the believers in what is written on notes, and the bold guessers lack.

To give a sense of the intuitive appeal of this approach, think of what happens when you summon up a mental image of an event described in a passage of prose. To make it image-like, you *have* to fill in the gaps; you have to include a red shirt kicking the winning goal from some part of the football field with some given trajectory, you have to make the goal scorer some putative size or other, you have to locate the goal somewhere, and so on and so forth. Much can be left indeterminate but you have to put in lots more detail than is delivered in the passage of prose. Also, you need to create a representation that represents inextricably. The 'part' that delivers the size of the scorer is also the 'part' that delivers the putative location of the scorer and the colour of the shirt. And so on. To the extent that you succeed, you create a state with a phenomenology.

Back to Mary on her release

So what is the before and after story about Mary? If feel is a matter of immediacy, inextricability, and richness of representational content, and the right kind of functional role, the difference is that, after her release, Mary has representational states with all those properties. If she makes the mistake of conflating intensional properties with instantiated properties, she will think that she has learnt something new about how things are, but she'll be wrong. Rather, she is in a new kind of representational state from those she was in before. And what is it to know what it is like to be in that kind of state? Presumably, it is to be able to recognise, remember and imagine the state. Once we turn our back on the idea that there is a new property with which she is directly acquainted, knowing what it is like to sense red can only be something about the new kind of representational state she is in, and the obvious candidates for that 'something about' are her ability to recognise, imagine and remember the state. Those who resist accounts in terms of ability acquisition tend to say things like 'Mary acquires a new piece of propositional knowledge, namely, that seeing red is like *this*' but for the representationalist there is nothing suitable to be the referent of the demonstrative.

We have ended up agreeing with Laurence Nemirow and David Lewis on what happens to Mary on her release.³² But, for the life of me, I cannot see how we could have known they were right without going via representationalism.

¹ Over the years I have received a large number of papers, letters and e mails seeking to convince me of the error of my old ways. Much of what I say below was absorbed from, or was a response in one form or another to, this material but I am now unsure who deserves credit for exactly what. More recently I am indebted to discussions of various presentations of 'Representation and Experience' in *Representation in Mind: New Approaches to Mental Representation*, H. Clapin, P. Slezaek and P. Staines (eds.) (Wesport: Praeger, to appear 2002).

² See, e. g., Frank Jackson, 'Epiphenomenal Qualia', *Philosophical Quarterly* **32** (1982), 127–36. The argument has a long history in one form or another. For an outline version

drawn to my attention recently, see J. W. Dunne, *An Experiment with Time* (London: Faber and Faber, 1927), 13–14.

³ This claim is common enough but it has been disputed on the basis of a Twin Earth argument. See Ned Block and Robert Stalnaker, 'Conceptual Analysis, Dualism and the Explanatory Gap', *Philosophical Review* **108** (1999), 1 – 46. For a response, see David J. Chalmers and Frank Jackson, 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* **110** (2001), 315 – 61.

⁴ See, e. g., Block and Stalnaker, *op. cit.*, but this is but one example among many.

⁵ See, e. g., Frank Jackson, Critical Notice of Susan Hurley, *Natural Reasons*, *Australasian Journal of Philosophy* **70** (1992), 475–87, and *From Metaphysics to Ethics: A Defence of Conceptual Analysis* (Oxford: Clarendon Press, 1998).

⁶ Most recently by Block and Stalnaker, 'Conceptual Analysis, Dualism and the Explanatory Gap', *op. cit.* For a fuller development of the reply in the text, see Frank Jackson, 'From H₂O to Water: the Relevance to *A Priori* Passage', *Real Metaphysics*, papers for D.H. Mellor, Hallvard Lillehammer, *et al.* (eds.) (London: Routledge, to appear 2002). Many once held, and some still hold, that the first premise, suitably fleshed out, is necessarily true as well as *a priori*. Nothing here turns on this issue. Incidentally, I am following the philosopher's lazy practice of simplifying the science.

⁷ See David J. Chalmers and Frank Jackson, 'Conceptual Analysis and Reductive Explanation', *op. cit.*

⁸ See, for example, Alex Byrne, 'Cosmic Hermeneutics', *Philosophical Perspectives*, **13** (1999), 347 – 83.

⁹ In my view, the illusion also fuels the modal intuitions encapsulated in the zombie, absent *qualia*, inverted *qualia* etc arguments, but I do not argue that here (though it may be clear how the argument would go).

¹⁰ G. E. Moore, 'The Refutation of Idealism', *Philosophical Studies* (London: Routledge and Kegan Paul, 1922), 1 – 30.

¹¹ David Hume, *Treatise of Human Nature*, bk. I, pt. IV, sec. VI.

¹² See, e. g., Gilbert Harman, 'The Intrinsic Quality of Experience', *Philosophical Perspectives* **4** (1990), 31 – 52.

¹³ These properties include the usual suspects like extension, colour and shape but I see no reason not to include, e. g., being an hydrometer. We can see something *as* an hydrometer. The difference between, *e.g.*, being extended and being an hydrometer is that you cannot see something without seeing it as extended whereas you can see something without seeing it as an hydrometer.

¹⁴ See, e. g., Frank Jackson, *Perception* (Cambridge: Cambridge University Press, 1977).

¹⁵ For a recent view of this kind, see John Foster, *The Nature of Perception* (Oxford: Oxford University Press, 2000), part three.

¹⁶ E. g., recently by Michael Tye, *Consciousness, Color, and Content* (Cambridge, Mass.: M.I.T. Press, 2000); see also Alex Byrne, ‘Intentionalism Defended’, *Philosophical Review* **110** (2001), 199 – 240. I should, perhaps, footnote what I think should be said about one example. The very same shape may have a different visual appearance depending on its putative orientation with respect to oneself. This in itself is no problem for representationalism, as orientation is part of how things are represented to be. However, as Christopher Peacocke points out, e. g., in ‘Scenarios, Concepts and Perception’ in Tim Crane (ed.) *The Contents of Experience* (Cambridge: Cambridge University Press, 1992), 105 – 135, seeing something as a regular diamond and as a square on its side need differ neither in putative shape nor orientation, and yet differ experientially. However, when this happens, one figure is being represented to be symmetrical about a line through its corners and the other about a line parallel to its sides.

¹⁷ How things are being represented to be need not be determinate. My experience may represent that something is a roundish shape without representing that it is any particular shape—the experience represents that there is some precise shape it has but there is no precise shape that the experience represents it to have. Indeed, it is arguable that all experience has some degree or other of indeterminacy about it. The same goes for maps and most sentences, of course.

¹⁸ I am indebted here to a discussion with Ned Block but he will not approve of my conclusion.

¹⁹ See, e. g., Michael Tye, *Consciousness, Color, and Content* (Cambridge, Mass.: M.I.T. Press, 2000) and *Ten Problems of Consciousness* (Cambridge, Mass.: M.I.T. Press, 1995). Tye's suggestion is *not* that the whole story about where the feel comes from lies in sensory states having nonconceptual content. But it is a key part of the story.

²⁰ As has been widely recognised, most recently in Richard G. Heck Jr., 'Nonconceptual Content and the Space of Reasons', *Philosophical Review* **109** (2000), 483 – 523; see also Tim Crane, 'The Nonconceptual Content of Experience' in Tim Crane, (ed.), *The Contents of Experience*, op. cit., 136 – 157.

²¹ I think it is the way Tye wants to go but I am unsure. But let me say that here, and in the immediately following, I draw on helpful if unresolved discussions with him.

²² For recent example, Richard G. Heck Jr., 'Nonconceptual Content and the Space of Reasons', op. cit. He is affirming it as an agreed view.

²³ I am here agreeing with Tim Crane, 'The Nonconceptual Content of Experience' op. cit., p. 140, but he would not, I think, agree with the use I make of the point on which we agree.

²⁴ Alex Byrne, 'Consciousness and Higher-Order Thoughts', *Philosophical Studies* **86** (1997), 103 – 129, see p. 117.

²⁵ Some argue that the two understandings are connected as follows: the reason for holding that belief contents are special in containing, in some sense, the relevant

concepts is that having a belief is special in requiring that one has the relevant concepts.

²⁶ As Christopher Peacocke puts it in *Sense and Content* (Oxford: Oxford University Press, 1983), 7 '[experience] can hardly present the world as being [a certain] way if the subject is incapable of appreciating what that way is'. Peacocke no longer holds this view.

²⁷ Michael Tye, *Ten Problems of Consciousness*, op. cit., p. 139, suggests that the key point is that to believe that something is *F* requires having a stored memory representation of *F* whereas to experience it as *F* does not. Thus, belief requires possession of the concept *F* in a way that experience does not. But one can believe that something is *F* for the very first time, and if the point is merely that one's system needs to have already in place the capacity to categorise something as *F*, that is equally plausible for both belief and experience. Christopher Peacocke, 'Analogue Content', *Proceedings of the Aristotelian Society*, supp. vol. **15** (1986), 1–17, points out that when we enter a room full of abstract sculptures, we perceive things as having particular shapes but need not have 'in *advance* concepts of these particular shapes' (p. 15, my emphasis). This is true but does not show that we do not have the concepts at the time we see the things as having the shapes.

²⁸ The talk of tagging the shade should not be understood on the model of a demonstration. According to representationalism, there need be no instance of the colour shade to be demonstrated.

²⁹ What drives the idea that the lack of words implies a lack of concepts sometimes

seems to be the modal claim that it is impossible to have words for all the shapes and colours we represent in experience, together with the plausible thesis that if I have the concept of, e.g., a certain shape, it must be possible for me to have word for it.

However, although it is impossible for me to have a word for every shape I discriminate; for any shape I discriminate, it is possible that I have word for it.

³⁰ I am here following David Armstrong but he should not be held responsible for the details.

³¹ D. M. Armstrong, *Perception and the Physical World* (London: Routledge and Kegan Paul, 1961), p. 128.

³² Laurence Nemirow, review of T. Nagel, *Mortal Questions*, *Philosophical Review* **89** (1980), 475 – 6, and 'Physicalism and the Cognitive Role of Acquaintance', *Mind and Cognition*, W. G. Lycan (ed.) (Oxford: Basil Blackwell, 1990), 490 – 99; David Lewis, 'What Experience Teaches', *Mind and Cognition*, op. cit., 499 – 19.