

**DOES THAT WHICH MAKES THE SENSATION OF BLUE A MENTAL FACT
ESCAPE US?**

BY JOHN CAMPBELL

**Forthcoming in Derek Brown and Fiona MacPherson (eds.), *The Routledge
Handbook of Philosophy of Colour* (Routledge, in press, expected 2016)**

DOES THAT WHICH MAKES THE SENSATION OF BLUE A MENTAL FACT ESCAPE US?

1. The Presence of Redness in Color Experience

So far as I know, the first clear statement of a broadly relationalist, primitivist view of color experience is in G.E. Moore's 'Refutation of Idealism'. The picture is 'relationalist' in that visual experience is thought of as a relation between the perceiver and the environment. And it's 'primitivist' in that the qualitative color, blueness or whatever, is taken at face value, as a characteristic of the external object, rather than being analyzed as a tendency of the thing to produce sensations or any other aspect of experience, and without any presumption that the qualitative color can be analyzed in physicalistic terms. The Moore passage is famous, but its significance is not usually appreciated:

that which makes the sensation of blue a mental fact seems to escape us: it seems, if I may use a metaphor, to be transparent – we look through it and see nothing but the blue.

(Moore 1903, 446)

If the qualitative character of blueness were itself mentalistic – a *quale* – then 'that which makes the sensation of blue a mental fact' could hardly be said to 'escape us'. Moore's

point here is that the qualitative blueness of the blue thing is not itself mentalistic; it is an external characteristic of the object out there. The experience itself is to be thought of as a matter of the perceiver standing in an experiential relation to that external blueness. This external blueness is what consequently constitutes the qualitative character of the experience.

This relational, primitivist picture of color experience obviously captures the way in which the color red itself is there in a visual experience of redness. The presence of the color red in a visual experience of redness is hard to explain on any view that takes your color experience to be a matter of you representing the colors in experience. Sometimes representationalists say that in visual experience, we have not representations but ‘perceptual presentations’ of the colors. I once heard someone say that in vision, we think about the colors, ‘*in color*’.

It seems sometimes to be supposed that the representationalist can capture the sense in which redness is ‘there’ in the experience by acknowledging that visual experience is ‘transparent’, in Moore’s term, in some sense that the representationalist can acknowledge. The idea is that when representing a situation, your attention is often focused not at all on the representation itself, but on the situation represented. If you and I are having a heated argument about whose turn it is to cook dinner, for example, our attention may be completely focused on the topic, and not at all on the words we’re using. Vision, the representationalist suggests, is simply a particularly striking form of this general phenomenon. Try to attend to your visual experience, and you end up focusing on the scene before you. Perhaps, it’s sometimes suggested, in vision it’s simply

impossible to attend to anything but the situation represented, you can't attend to the visual representation itself (cf., e.g., Tye 2013).

Whatever the representationalist should say about the modality here (is it really completely 'impossible', or just something we usually don't do, or what?), it should be evident that this kind of idea can't begin to explain the sense in which redness itself is 'there' in a visual experience of redness. Suppose that you and I are having an argument about the battle of Gallipoli. Our attention may be focused entirely on the situation represented; we are not thinking at all about the words we are using, or how we are talking about the people in question. In this case it's quite natural to say that the words are 'diaphanous', our attention falls through them and onto the things we are thinking about. But 'transparency' in this sense is not enough to make the situation at Gallipoli 'present' to us in the way in which colors are present in color experience. The sense in which redness is 'there' in an experience of red can't be explained in terms merely of the transparency of visual representation.

Traditionally a natural reaction at this point has been to appeal to some non-representational aspect of visual experience, such as a 'sensation', in order to explain the sense in which redness is there in the experience. Representations merely tell you what is out there; sensations make the scene come alive, pump color into it. One idea is that color terms apply in the first instance to those sensations, and only derivatively, if at all, to physical objects. The sense in which redness is there in vision is that vision involves not only representations, but sensations which are, literally, red. Whether external objects can also be said to be 'red' is a further question, sometimes thought to have been decided negatively by science. The trouble with this idea is that our ordinary color

concepts quite obviously apply in the first instance to physical objects. To grasp color concepts one has to have some at least implicit understanding of color constancy – that objects typically aren't changing their colors just because the conditions of illumination change, for example – and this grasp of constancy makes no sense for the case of sensations, which don't have 'conditions of illumination' (for further discussion of color constancy, see Smith 2002, Brown 2014 and 'Colour Constancy', this volume).

You might therefore retreat to a more oblique use of sensations. You might say that redness is 'there' in ordinary visual experience in the sense that not only is redness represented, there is a suitable sensation accompanying that representation. But just having some sensation or other could not amount to redness itself being present in the visual experience, no matter how often that sensation accompanies the representation of redness.

Another line of thought for representationalism begins with the difficulty that the usual naturalistic accounts of perceptual representation have with color experience. What is it for a cognitive system to represent redness? We need the system to be causally sensitive specifically to the presence or absence of redness, and for that to be the point of the system. But of course it's not only perceptual systems that could 'represent' redness in this sense. Any kind of linguistic system for talking about redness will represent redness in that sense. So the representation of redness, in this sense, doesn't bring with it the presence of redness itself.

You might then suggest that we need some further notion of 'phenomenal' representation, which will bring with the representation of redness the perceptual presence of redness. It's easy to see the pressure that leads the theorist to make this kind

of claim. It's less easy to see how to make sense of the claim. How can the mere representation of a property demand the presence of the property itself?

One way of trying to make sense of this is to say that the perceptual representation of redness represents the *essence* of redness; that's the sense in which redness is 'present' in perception but not in other ways of representing redness. The trouble with this is to find an interpretation of 'essence' on which this kind of claim could be both true and a plausible description of perceptual presence. What is the essence of a color? A natural thought is that this has something to do with the geometry of color – the location of particular shades in the color solid. So you might think that the essence of orange is that it's a binary color located between red and yellow. But this description of the essences of the colors as having to do with their structural characteristics simply misses the key way in which redness is there in color experience. It's not merely the *structure* of the colors that is present in color experience. It's the absolute values of the colors themselves – the blueness of the blue thing, for example.

What you want to do is point to the qualitative character you're encountering in experience and say '*that*, that is what's essential to blueness'. So you might say: the sense in which blueness is present in experience is that the qualitative character that is essential to blueness is present in experience. But now we simply have arrived at the relational view. The relational view is that the experience itself is a relation to the blueness of the external object. That external blueness is the only thing you can be pointing to when you say, '*that*, that is what's essential to blueness'. It's the external qualitative character that matters.

2. The Physical Bases of Color and Color Experience

I've been arguing that our only conception of 'the redness of a red experience', or 'phenomenal redness', is that the experience is a relation one stands in to that mind-independent property of redness. On a primitivist conception, colors are mind-independent characteristics of the objects and liquids and so on around us. Our understanding of one another's color experiences depends on our knowledge of the colors, so conceived.

Color, conceived in this way, is an indispensable element in our imaginative understanding of one another. Imaginative understanding has its own epistemic authority. We know of the existence and nature of one another's conscious lives on the basis of our imaginative understanding of each other. If we didn't have an imaginative understanding of one another, physical science couldn't provide us with knowledge of one another's conscious states. That was the point about Nagel's bat.

Even though imaginative understanding has its own epistemic authority, we can still ask whether the mind-independent colors we think about are grounded in the physical world. Here we hit problems. Given any candidate physical basis for color, you can imagine the presence of the color without the presence of that physical basis. You are more certain of the presence of the color than you are of the presence of any physical basis. And the color seems to have a certain unity and simplicity that the physical basis altogether lacks.

As Byrne (2006) notes, these arguments are counterparts of Descartes's arguments for the distinctness of mind and body. We do not seem to gain any traction in our understanding of the relation between the macroscopic world and the underlying physical reality by trying to explain color in terms of some mentalistic construct and then explaining that in terms of the brain. Moreover, as we have seen, we do not seem to have a good way of explaining the required mentalistic construct – one that would acknowledge the sense in which color is present in perceptual experience.

You might say, though, that the structure of the qualitative colors has a physical counterpart in the structure of the neural assemblies involved in color vision, but has no physical counterpart in the structure of the physical environment of the observer. You might argue that it follows from this that to explain how the qualitative character of color is grounded in physics, we need a two-stage procedure:

- (1) Explain the qualitative characters of the colors of objects in terms of the qualitative character of color experience, and then,
- (2) Explain the qualitative character of color experience in terms of the structure of neural processes involved in color vision.

Is this argument any good? At the most abstract level we can see right away that it isn't. We can have a whole high-level explanatory structure in which there are mind-independent colors grounding our color experiences, and yet the only physical systems that have a similar structure, identified now by its significance at the physical level, are

neural, or defined by their relations to the neural. The whole high-level explanatory system could still be grounded in the physical, without our needing to postulate some notion of the ‘qualitative character of color experience’, identifiable independently of the mind-independent colors themselves. But let’s look further at the argument.

Let’s accept for the moment that the colors of the objects and liquids and flames and so on do have a determinate structure, and that there is no corresponding structure in any of the physical characteristics of the objects and liquids and flames and so on. And let’s also accept, for the moment, that there is a correspondence between the structure of the colors and the structures of neural assemblies.

On a primitivist view of color, the colors of objects and so on supervene on the physical characteristics of the environment, but may not have any explanatory or reductive grounding in those physical characteristics. The argument that reduction may not be possible is the one already rehearsed: that we can readily conceive of color and physical composition coming apart, and so on. The structure of color is, therefore, the structure of a high-level set of absolute properties that cannot be reduced to the physical. There is, therefore, no *a priori* reason to suppose that this structure must be reducible to the physical.

There is, of course, nothing particularly remarkable about the idea that there may be high-level structures without any echo at the level of physics. Consider, for example, the dynamics of a credit squeeze. So far as the high-level economic properties go, there are quantitative models for what is happening: the failure of confidence, the liquidity crises for businesses, and so on. Now all this economic structure presumably supervenes on what is going at the level of gluons. But that’s not to say that we should expect there

to be quantum-mechanical structures, with independently identifiable quantum-mechanical significance, that correspond to the structure of a credit crunch.

Consider now what has to happen for there to be perceivers who can see those high-level properties, the colors. These perceivers must have perceptual systems that are causally responsive to those high-level structures. Now the scientific study of human perceptual systems generally assumes them to be describable at 3 levels. There is the level of computation, at which the task of the perceptual system is specified. There is the level of the algorithm, at which we characterize what strategies the system uses to that task. And there is the level of implementation, at which we say how these computations are implemented biologically (Marr 1982, cf. Griffiths et. al. 2015). If, at the levels of computation and algorithm, we are describing representations that are sensitive to the high-level structure of the colors, then anything recognizable as a biological implementation of those representations will have to have a corresponding biological structure.

In other words, on the most straightforward reading of color primitivism, the natural expectation would be that on this view, there may be no physical structures in the environment that correspond to the structure of the colors, but that we would expect there to be biological structures that echo the structure of the colors. Suppose now that a philosopher claims that this is exactly the situation that empirical science discovers (cf. Pautz 2014). It is difficult to see why this should be thought of as a threat to primitivism, rather than being a confirmation of it.

There is, however, a rather larger and vaguer way of interpreting this line of argument, one that does not claim that there is some direct contradiction between the

empirical scenario envisaged and color primitivism. Rather, the argument might be interpreted as an attempt to find considerations motivating one rather than another direction of research in attempting to ground color phenomena in the physical. The idea would be that because of the difficulty of grounding the structures of the colors in the structures of the physical environment, the only path to take is to ground colors in color experiences, and to ground color experience neurally.

This argument limps at every step. The idea is that we should explain the structures of the colors of objects and so on in the structure of color qualia, and then ground the color qualia neurally. There are two problems with the proposed direction of research:

- (1) We have no idea what we are talking about when we talk about ‘color qualia’. I will amplify on this point in the next section, but we have already seen the basic problems. We can’t explain color qualia in representational terms, because doing so misses out the ‘presence’ of the color itself in color experience. When we try to add in the color, saying that here we have a ‘special type’ of representation, we simply bring in the external color itself. If we try to appeal to ‘color sensations’, the basic problem we hit is that we have no way of saying what the various types of sensations are that we are talking about. We have no vocabulary to describe them.
- (2) The idea is that to find a physicalist reduction of color, all we have to do is to abandon primitivism and make sense of color qualia; then all that then stands in

the way of a reduction of color is resolution of the mind-body problem (framed in terms of 'qualia'). This may be part of a general program, to reduce the relation between the medium-sized world generally and the world of physics to the mind-body problem. But this is not a promising program; all we have done is make the mind-body problem insoluble.

To repeat, this is not to say that neural structure may not ultimately have a role to play in explaining the physical basis of the colors, but that the shuffle through an internalist construct, the 'color qualia', merely leads to confusion: the construct cannot be properly explained.

3. Color Qualia

Let us look a little further at how people try to explain what they mean when they talk of 'color sensations', or 'color qualia'. In a recent article, Ned Block writes:

Are phenomenological characters of perception—e.g. what it is like to experience redness or roundness—philosophically reducible to the redness or roundness of the objects one sees or to representation of redness or roundness? If there is no such reduction, then there can be said to be mental paint.

(Block 2010, 23-24)

This seems to be his official explanation of what he means by ‘mental paint’. Block is evidently struck by the thought that something more might be wanted, so he provides the following comment for the puzzled reader:

I am not assuming that if there is mental paint, it is non-relational (“intrinsic”) or has no representational aspect. Since I favor physicalism, I allow that mental paint may be a relational neural property. To avoid misunderstanding: I do not claim that there is anything red or round in the head when one veridically sees a red or round thing in the world as when red pigment in a painting represents a red barn.

(2010, p. 56 footnote 2)

And that is it. So far as I can see, at any rate, that’s all we get by the way of a positive description of what ‘mental paint’ is supposed to be. In all his voluminous writings on this topic, Block nowhere, so far as I know, provides any further substantial detail as to what mental paint is. There is, for example, no attempt to say what vocabulary you should use to specify particular types of mental paint.

I think that the most charitable reading here is that Block is using ‘mental paint’ as a term for a theoretical construct, whose further characterization has yet to be given. Block will introduce examples intended to show that no relational or representationalist theory can give a full description of color perception. ‘Mental paint’ is simply his name for whatever it is that is missing. A parallel might be the use that physicists make of

terms like ‘dark energy’ or ‘dark matter’. When it’s discovered that the rate of expansion of the universe is greater than current theories can explain, ‘dark energy’ is introduced as a name for whatever it is that is doing that. Someone using the term ‘dark energy’ typically isn’t under any illusion that they have to hand a full characterization of the explanatory role that dark energy is playing. Still less are they under the illusion that they have an explicit characterization of what dark energy is. Still, the term ‘dark energy’ has some use, as pointing to some phenomenon that needs further explanation. Just so for Block’s ‘mental paint’. Block intends to point to phenomena of color experience that current theories cannot explain; ‘mental paint’ is simply his name for whatever it is that is missing.

The trouble is that when we are talking about color experience, it is hard to see how there can be any role for hidden aspects of the thing that can only be discovered by complicated thought-experiments or work in experimental psychology. Of course, students of the mind have always been fascinated by the idea of arcane levels of psychology not known to the ordinary person, from Kant’s transcendental psychology through Freud’s psychodynamics. Color qualia do seem sometimes to be thought of as an arcane level of color experience, unsuspected by the ordinary perceiver, who sees only the colors of objects. But the whole point about color experience is that it has to do with what’s on the surface; it can’t be a home for arcana. The trouble is that ‘mental paint’ is supposed to figure in a characterization of what color experience is like. How could that be consistent with taking mental paint to be a theoretical postulate like ‘dark energy’ whose characterization we have yet to discover?

In an influential discussion, Christopher Peacocke gave one of very few serious attempts to provide an explicit vocabulary for talking about color sensations (and subsequently did more than any other philosopher to show how this kind of vocabulary might be put to work):

It will help at this point if we introduce a simple piece of notation. If a particular experience e has the familiar sensational property which in normal circumstances is produced by a white object (such as a tilted plate) which would be precisely obscured by an opaque elliptical region (r , say) of the imagined interposed plane, let us express this fact in the notation elliptical' (r, e) and white' (r, e).

(Peacocke 1983, 20)

I think the key phrase here is 'familiar sensational property'. Peacocke is taking it that we all already know perfectly well the phenomena he is talking about when he talks about 'red' sensations'. That is why he does not think that the concepts need anything much in the way of explanation. But are we 'familiar' with sensational color properties in vision? Recall Moore's remark 'that which makes the sensation of blue a mental fact seems to escape us'. All that experience gives us knowledge of, on Moore's view, is the blueness of the external object. How could you suppose that there is such a thing as the 'familiar sensational property' of being 'blue', since what you encounter in experience is only the blueness of the external object? And in fact Moore is surely right about this: we

don't encounter in experience any color property other than the external property of things in the environment.

The dilemma you face in trying to explain the idea of 'color qualia' is that you either (a) assume that in ordinary vision, when you see 'the blueness of the blue thing', you are merely encountering some aspect of your subjective life, not an aspect of your surroundings, in which case you deny the Moorean transparency of experience, and with it, the ability to recognize vision as giving you knowledge of what your surroundings are like, or (b) you take it that the very idea of 'color qualia' is a theoretical construct, a posit to be characterized by describing its theoretical role. You might then try to give a fully explicit characterization of this theoretical role. This would then be comprehensible to Martian anthropologists who have no color experience. But the whole idea of color qualia was to describe 'what it is like' to have color experience in terms of them, so how can we think that someone could have a full understanding of color qualia without having any conception of what color experience is like?

You might argue that in ordinary color experience we are aware of two different phenomena, the redness out there and the 'red'-ness in the mind'. Thus Peacocke says, 'In the case of the sensational properties, the experienced properties are, in the visual case, properties and relations of the visual field and its parts.' (Peacocke 2008). It is very hard to know how this idea is supposed to work. Does red'ness, as a characteristic of the visual field, exhibit color constancy (that is, can something continue to be manifestly red' through changes in the illuminant)? Does red'ness exhibit simultaneous color contrast (that is, does the apparent red'ness or not of a region depend on the colors' of neighboring regions of space? If so, then 'red'' seems to be none other than the

ordinary term 'red', with the prime indicating merely that it is being applied to some delimited region of space. Most of these ascriptions will then be false, since regions of space can't in general be said to have colors. The other possibility is that red'-ness does not exhibit color constancy or simultaneous color contrast, in which case I think we have as yet simply no idea what it means, although we're assured that it stands for a phenomenon that we're 'familiar' with. But ordinary visual experience seems to familiarize us only with the colors of the external objects.

The problems that we face in trying to interpret Block and Peacocke are exactly those that we should expect, if color experience is in fact a relation between the perceiver and an external color. Anyone trying to find some middle aspect of color experience, 'the phenomenal color', is then caught between regarding this thing as a purely theoretical postulate, like dark matter, or claiming that it's something we encounter in everyday experience, when in fact we encounter no such thing. A third option, of course, is to oscillate between these two positions.

I think the right reaction is to reflect on what materials we have for talking about color and color experience that clearly do make sense. And here there are two types of vocabulary:

- (1) The words we have in ordinary natural language for the colors of things – the redness of a red pepper, and so on.
- (2) The terms that vision scientists use for describing the information-processing that takes place when we see colored things.

The relational, primitivist view of color experience tries to characterize color experience simply in terms of these vocabularies. It takes the relation 'X experiences color Y' to be a generic relation that can hold between any perceiver and any color in the environment. Or this austere position can be augmented: we can take it that what we have is a three-place relation, 'X experiences color Y from standpoint Z', where the characterization of the standpoint might include, for example, specification of the conditions of illumination. Our understanding of color experience is achieved by the exercise of the imagination, and in exercising imagination we take up the standpoint of the perceiver. The relation of experience is itself a theoretical construct; as Moore observed, it is this relation of experience that makes the sensation of blue a mental fact (not some 'blue' *qualie*), and in ordinary experience this relation 'escapes us', it cannot be ostensively explained. Our grasp of this construct is implicit in the way we use imagination to understand how other people are experiencing the colors around them.

From the perspective of anyone who thinks that color experience should be characterized as a three-place relation in this way, between the perceiver, the external color, and the standpoint of the perceiver, the problem with thinking in terms of an 'inner' realm of qualitative color qualia is that (a) the external, qualitative color of the object has been regarded as an intrinsically mentalistic phenomenon, when in fact we have no way of identifying any such thing, and (b) the fact that we are dealing with a three-place relation has been misconstrued as implying that we are dealing with a two-place relation between the perceiver and the external phenomenon, mediated by the constituents of an inner realm of qualia. The problem here was already stated by Quine:

‘immediate experience simply will not, of itself, cohere as an autonomous domain’ (1960, p. 2).

As we saw earlier, people who think that visual experience is representational may hold that there is a concept of ‘phenomenal intentionality’ we can use to characterize color experience. This will be like the ordinary notion of ‘representing something as red’, except that the ‘phenomenal representation’ requires the actual presence of redness itself.

The trouble is that there seems in principle to be no way of explaining the required conception of ‘phenomenal intentionality’. ‘Intentionality’ itself is a fairly definite phenomenon. There are ordinary beliefs, which on the face of it illustrate intentionality, and goal-directed behaviors. But merely saying that color experience is ‘intentional’ would leave out the striking difference between visual experience of the colors and having beliefs about colors. The difference is that in visually experiencing redness, redness itself is present, and that isn’t so for having a belief about redness.

The ‘rich’ intentionalist tries to acknowledge this point by saying that we have here a particular variety of intentionality, ‘phenomenal intentionality’. For a statement of this idea, consider Chalmers’ approach:

The phenomenal character of a perceptual experience is what it is like to have that experience

A representational content of a perceptual experience is a condition of satisfaction of that experience

A *phenomenal content* of a perceptual experience is a representational content that is determined by the experience's phenomenal character.

(2006, 50)

The key term here is 'what it is like'. On one reading, perhaps not the intended reading, of this approach, 'what it is like' to have the experience is to be characterized in terms of qualia.

- (1) 'What it is like' to have an experience is a matter of the qualia constituting that experience. The phenomenal character of an experience of yellow, for example, will be characterized by a yellow quale.

There will then be the question how this internal qualitative character determines a representational content, and a natural suggestion is this:

- (2) The representational content determined by that quale will be to the effect that some external object has an external property of yellowness. That external property of yellowness will be a resemblance of the qualitative character of yellowness possessed by qualia.

This idea that external objects might be represented as having properties that resemble the yellowness and so on of *qualia* then gives us a reading of what Chalmers means by 'Edenic' representational contents.

There are two problems with this conception of ‘phenomenal intentionality’. One is that the talk of ‘inner yellowness’, yellowness as a characteristic of qualia, makes no sense. We can’t explain it ostensively, because in ordinary visual experience we encounter only the yellowness of the external object. And we can’t explain it as a theoretical construct, like ‘dark matter’, because we’re trying to use it to characterize visual experience itself. So the talk of yellowness as a characteristic of qualia can’t be sustained. The second problem is that the talk of ‘resemblance’ between the external objects and the qualia makes no sense. Even if we did have some conception of what it is for a purely mental item to be yellow, that would not of itself mean that we understand what it is for an ordinary physical object to resemble the mental item in point of color.

An alternative reading is to set aside the talk of qualia, and interpret the talk of ‘what it is like’ not in terms of qualia, but in terms of our imaginative understanding of one another. Here we could use the idea of experience as a three-place relation between the perceiver, a standpoint and the scene observed, as sketched about. On this reading, ‘what it is like’ to perceive a scene is specified by specifying the scene observed, the observer, and the point of view from which the observer is perceiving the scene. So you would specify a color experience by saying which colors are there to be seen, who the observer is, and the standpoint from which the observer is seeing those colors. There is, of course, nothing representational about this way of saying what the experience is like for the observer. It would indeed, as Chalmers says, be possible to use that three-place relation to ‘determine’ any of a wide variety of representational contents, depending on which ‘determination’ relation is arbitrarily chosen by the theorist. But the specification of a representational content here will be idle in giving a description of the experience, it

adds nothing to what is available at the level of imaginative understanding. We could eliminate all the talk about ‘representation’ and still have a full grasp of color experience.

4. Imagining *De Re*

There is a line of thought that can make the idea of a relational view of color experience seem not just problematic but utterly incomprehensible. In a famous paper, Thomas Nagel wrote:

At present we are completely unequipped to think about the subjective character of experience without relying on imagination - without taking up the point of view of the experiential subject.

(Nagel 2002, 224)

I think it is fair to say that Nagel’s comment here has not been effectively challenged in the voluminous subsequent literature. But if knowing about someone’s experiences of color requires imagining them ‘from the inside’, how can we so much as form the conception of experience as a relation between the experiencer and the qualitative environment? To address this, we have to look further at what it means to be imagining someone’s experiences of color.

Suppose you have committed a crime, let's say a murder. You have disposed of the body and are waiting to be interviewed by detectives. As they stand talking to you, you imagine how they are seeing the room. Can they see anything incriminating? Will they see the flashlight? What you are doing here is an exercise in imagining *de re*. That is, the exercise of the imagination constitutively depends on your perceptual knowledge of what is there in the room. You begin with your knowledge of the room, you can see quite well what's there and where it all is. You imagine *of that stuff* how it is seen by the detectives. Your imaginative understanding of their visual experiences constitutively depends on your knowledge of their surroundings. In contrast, what we might call 'imagining *de dicto*' is imagining someone's experiences in a way that does not constitutively depend on knowledge of their environment. It seems arguable that imagining *de dicto* depends on the ability to imagine *de re*. Imagining someone's visual experiences *de dicto* is a matter of imagining an environment for them; and then, within the context of that imaginative exercise, using one's capacity to imagine *de re* what the person is seeing. Imagining *de dicto* depends on imagining *de re* because imagining *de dicto* involves an exercise of the capacity to imagine *de re*.

To illustrate the point, recall Jackson's Mary, born in a black and white room, a brilliant scientist who knows all there is to know about the physics of color and the science of color vision (Jackson 2002). Jackson's question was, when Mary steps out into the world of color, will she learn anything? His own answer was that she will: she'll learn what it's like to have the experience of color. However, on the face of it, what Mary learns about in the first instance is not a psychological phenomenon at all. It's the surfaces of objects that she learns about: she learns what the colors of objects are. To

see this, suppose we tweak the example a little. It's often said that people with autism have difficulty with mentalistic concepts. Whatever the merits of this as an approach to autism, suppose that this kind of analysis applies to Mary. Suppose that, though intelligent, she has no concept at all of visual experience. Then when she steps into the world of color, she will not learn about any psychological phenomena. Will she then learn nothing at all? It seems evident that there is still a lot she will learn. She will learn about the colors of objects, even though she is learning nothing about visual experience.

This suggests that even when the original Mary is learning about visual experiences, she is able to do so only because she has first learned about a more basic phenomenon, the colors of the objects around her. She has first learned about the colors of the things around her. Now, *of* that phenomenon, she can imagine what it's like for people to see it. Her capacity to reflect on people's visual experiences of color is provided by her capacity to imagine, *de re*, of the colors she knows about, what it is like to see *them*. Imagining people's experiences of color requires first having knowledge of the colors themselves, and only then, in the context of that knowledge, imagining what it is like to see them. It is also true that you can imagine a group of colored objects surrounding a person, and then, in the context of that imaginative exercise, imagine the person's experiences of those colors. But this is plainly a derivative exercise of imagination *de re*. The basic case is that in which your imaginative understanding depends on your knowledge of the shared environment.

It is, of course, possible to argue that physics has shown that color, as a characteristic of the objects around us, simply does not exist. A full physical story about the objects around us does not involve ascribing colors to them. Nor does it involve

ascribing to them any physical characteristics from which their colors might be derived. Right now, I remark only that the same line of argument would show that consciousness, as a characteristic of you or I, simply does not exist. A full physical story about ordinary humans does not involve ascribing consciousness to them. Nor does it involve ascribing to them any physical characteristics from which their conscious states must be derived. Since we are conscious, there must, therefore, be something wrong with this argument against the existence of colors, conceived as characteristics of the objects around us.

We can now reformulate the idea that color experience should be thought of as a relation between the perceiver and the colors in the surroundings. If you think that all imagining is imagining *de dicto*, and the possibility of imagining *de re* does not even occur to you, then the idea of color experience as a relation will seem absurd. Someone's color experiences are what you can understand imaginatively; if imaginative understanding relates only to what is 'confined to the head', and can have no constitutive dependence on your knowledge of the surroundings of the other person, then color experience itself must be confined to the head, and cannot be a relation to the perceiver's surroundings. The idea of color experience as a relation to the surroundings is the idea of color experience as something that has to be understood by imagination *de re*.

Recall the remark from Moore: 'that which makes the sensation of blue a mental fact seems to escape us: it seems, if I may use a metaphor, to be transparent – we look through it and see nothing but the blue' (Moore 1903, 446). As we saw, the key point here is that the blueness of the blue thing – that with which we are unmistakably confronted in color experience – is not, on this account, a mental phenomenon, it is not to be thought of as a *quale* produced by the external object, for example. The blueness is

out there, on the object. The experience of blueness is a *relation* to that external phenomenon. We understand what this experience is by imagining *de re*. Nonetheless, the relation itself, the relation of *experiencing*, which is what makes it the case that we have a mental fact here, is not itself something that we confront, either in the initial perception or in having an imaginative understanding of it.

5. The Inverted Spectrum

The possibility of spectrum inversion is usually taken as a datum in discussions of color and color experience (for discussion, see ‘Spectrum Inversion’, this volume). What is certainly a datum is that you can imagine a world in which all the objects have different colors than they do in this one. You can imagine a world in which the fire engines are blue and the sky is red and so on. And you can imagine someone experiencing those colors.

What is meant by spectrum inversion, however, is usually the possibility that the colors in the environment are held constant, but that each of us experiences them differently. It is not at all obvious that this is possible. The use that’s being made of the imagination here is relatively complex, and it’s not at all obvious that this is a correct description of any imaginative exercise one is likely to carry out.

Why does it matter? Well, one way you might argue against color primitivism is to say that (a) people’s color qualia might all be swapped around, and then (b) there would be no saying who was experiencing the colors correctly, but (c) color primitivism

implies that there would have to be just one correct way to experience the colors (cf., e.g., Chalmers 2006).

This line of argument does not give due weight to the point that imagining color experience *de re* is more fundamental than is imagining color experience *de dicto*. If you think that imagining other people's color experiences is constitutively independent of your perceptual knowledge of their environment, then it makes perfect sense to think of varying everyone's color experiences while holding constant the physics of the external environment. And you might well then face a difficult problem in saying which, if any, of these internalistic experiences resembles the external environment.

It is quite different if you are looking at the colors in your environment and imagining what other people are seeing when looking at *those colors*. Suppose you're a painter, working on a canvas. The coloration overall seems to you a bit heavy and flat, and you're working to lighten it a little and give it some more drama. You're imagining what other people will see when looking at *this*. Now if you've been reading about the inverted spectrum you might try to imagine what your canvas would be like with all the colors reversed, and then, within the compass of that imaginative project, you could imagine what it would be like to see it from various positions and distances. That imaginative exercise is providing you with information about what it would be like to be seeing all the colors round the other way from you. But it's not as if you can now sensibly ask the question, 'Does my color experience provide knowledge of the colors of the canvas any better than my hypothetical subject's experiences provide knowledge of the colors that same canvas?'. .

Of course your own experiences are providing you with knowledge of the colors of the canvas before you. And of course this other subject, who you are imagining as seeing the colors on a quite different, color-reversed, merely hypothetical canvas, is not getting knowledge of the colors on the canvas before you. That's just built into the description of the two imaginative exercises. Once we give due weight to the *de re* character of imaginative understanding, the idea of holding the external environment constant while varying the subjective experiences that people are having of it simply has to go.

REFERENCES

Block, Ned. 2010. 'Attention and Mental Paint'. *Philosophical Issues*, 20, 23-63.

Brown, Derek H. 2014. 'Colour Constancy and Colour Layering'. *Philosophers' Imprint*, 14, 1-31.

Byrne, Alex. 2006. 'Color and the Mind-Body Problem'. *Dialectica* 60, 223-244.

Chalmers, David. 2006. 'Perception and the Fall From Eden'. In Tamar Gendler and John Hawthorne (eds.) *Perceptual Experience*. New York: Oxford University Press.

Griffiths, Thomas and Noah Goodman. 2015. 'Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic'. *Topics in Cognitive Science* 7, 217-229

Jackson, Frank. 2002. 'Epiphenomenal Qualia'. In David Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.

Marr, David. 1982. *Vision*. San Francisco: W.H. Freeman.

Moore, G.E. 1903. 'The Refutation of Idealism'. *Mind* 48, 433-453.

Nagel, Thomas. 2002. 'What is it Like to be a Bat?'. In David Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press.

Pautz, Adam. 2014. In R. Brown (ed.), *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience*, Studies in Brain and Mind 6. Dordrecht: Springer.

Peacocke, Christopher. 1983. *Sense and Content*. Oxford: Oxford University Press.

Peacocke, Christopher. 2008. 'Sensational Properties: Theses to Accept and Theses to Reject'. *Revue Internationale de Philosophie* 243, 7-24.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Smith, A. D. 2002. *The Problem of Perception*. Cambridge, Mass.: Harvard University Press.

Tye, Michael. 2013. 'Transparency, Qualia Realism and Representationalism.' *Philosophical Studies* 170, 39-57.

