

## **Physicalist Theories of Color**

**With J. David Velleman.**

### **THE PROBLEM OF COLOR REALISM**

The dispute between realists about color and anti-realists is actually a dispute about the nature of color properties. The disputants do not disagree over what material objects are like. Rather, they disagree over whether any of the uncontroversial facts about material objects—their powers to cause visual experiences, their dispositions to reflect incident light, their atomic makeup, and so on—amount to their having colors. The disagreement is thus about which properties colors are and, in particular, whether colors are any of the properties in a particular set that is acknowledged on both sides to exhaust the properties of material objects.

In a previous paper we discussed at length one attempt to identify colors with particular properties of material objects—namely, with their dispositions to cause visual experiences. Here we shall discuss a different and perhaps more influential version of realism, which says that the colors of material objects are microphysical properties of their surfaces.<sup>2</sup> We shall call this theory physicalism about color (physicalism, for short). In order to evaluate this theory, however, we shall first have to clarify some methodological issues. Our hope is that we can bring some further clarity to the question of color realism, whether or not we succeed in our critique of the physicalists' answer.

### **Metaphysics and Semantics**

To say that the question of color realism is really about the nature of color properties is not yet to define the question sufficiently. One is tempted to ask, Which are the properties whose nature is at issue?

Of course, the latter question may seem like an invitation to beg the former. For in order to say which properties are at issue in the debate about the nature of colors, one would have to say which properties colors are—which would seem to require settling the debate before defining it. How, indeed, can one ever debate the nature of a property? Until one knows which property is at issue, the debate cannot get started; but as soon as one knows which property is at issue, it would seem, the debate is over.

Well, not quite. One can pick out a property by means of a contingent fact about it. And one can thereby specify the property whose nature is to be debated without preempting the debate. Such indirect specifications are what motivate questions about the nature of properties. One knows or suspects that there is a property playing a particular role, say, or occupying a particular relation, and one wants to know which property it is, given that playing the role or occupying the relation isn't the property in question.

The role in which colors command attention, of course, is their role as the properties attributed to objects by a particular aspect of visual experience. They are the properties that objects appear to have when they look colored. What philosophers want to know is whether the properties that objects thus appear to have are among the ones that they are generally agreed to have in reality.

Yet if the question is whether some agreed-upon set of properties includes the ones that objects appear to have in looking colored, then it is partly a question about the content of visual appearances. When philosophers ask whether colors are real, they are asking whether any of the properties acknowledged to be real are the ones attributed to an object by the experience of its looking colored; and so they are asking, in part, which properties are represented in that experience—which is a question of its content.

### **What is Looking Colored?**

The foregoing attempt to define which properties are at issue in the question of color realism may seem viciously circular. For we identified colors as the properties that things appear to have when they look colored; and how can this description help to pick out the relevant properties? It specifies the properties in terms of their being represented in a particular kind of experience, but then it seems to specify the relevant kind of experience in terms of its representing those properties. Which properties objects appear to have in looking colored depends on what counts as looking colored, which would seem to depend, in turn, on which properties colors are—which is precisely what was to be defined.

This problem is not insuperable, however. The phrase "looks colored" and its determinate cousins—"looks red," "looks blue," and so forth—have a referential as well as an attributive use. That is, one learns to associate these phrases directly with visual experiences that are introspectively recognizable as similar in kind to paradigm instances. Paradigm cases of looking red fix a reference for the phrase "looks red," which then refers to all introspectively similar experiences. We can therefore speak of something's looking red and rely on the reader to know which kind of visual experience we mean, without our having to specify which property red is.<sup>3</sup> There is no circularity, then, in identifying red as the property attributed to objects by their looking red and, more generally, in identifying colors as the properties attributed to objects by their looking colored.<sup>4</sup>

### **Color Experience vs. Color Discourse**

One might think that the references we have stipulated here are simply the references that color terms have anyway, in ordinary discourse. Surely, words like "red" and "blue" are sensory terms, designed to report what is seen. One may therefore feel entitled to presuppose that the term "red," as used in ordinary discourse, already denotes the property that things appear to have when they look red.

Yet the validity of this presupposition may depend on the answer to the question of color realism. For whether the ordinary term "red" always expresses the property that things appear to have in looking red may depend on whether that appearance is veridical or illusory. Suppose that an error theory of color experience is correct, in that the property that things appear to have when they look red is a property that they do not (and perhaps could not) have. In that case, the meaning of "red" in ordinary discourse will be subject to conflicting pressures. The term may still be used to express the property that objects are seen as having when they look red. Yet statements calling objects red in that sense will be systematically false, even if such statements tend to be made, and to garner assent, in reference to objects that have some physical property in common. In the interest of saying what's true, rather than what merely appears true, speakers may then be inclined to shade the meaning of "red" toward denoting whatever property is distinctive of red-looking objects.<sup>5</sup> The pressure towards speaking the truth will thus conflict with the pressure towards reporting the testimony of vision. How the meaning of "red" will fare under these conflicting pressures is hard to predict; it may even break apart, yielding two senses of the term, one to express the content of color experience and another to denote the property tracked by color attributions.

We are not here proposing or defending such an account of color language. We are merely pointing it out as a possibility and suggesting that this possibility shouldn't be excluded at the outset of inquiry about color. To assume that color terms denote the properties represented in color experience is to assume that terms used to attribute those properties to objects wouldn't come under pressure from the systematic falsity of such attributions—something that may or may not turn out to be the case but shouldn't be assumed at the outset. One should begin as an agnostic about whether color terms ordinarily denote the properties that are represented in color experience. If they are to be used in a debate about those properties, their reference to them must be explicitly stipulated.

## **VERSIONS OF PHYSICALISM**

If physicalism is to settle the debate over color realism, it must be formulated as a thesis about the properties at issue in that debate. When the physicalist says that colors are microphysical properties, he must mean that microphysical properties are the ones attributed to objects by their looking colored. Otherwise, his claim will not succeed in attaching the uncontroversial reality of microphysical properties to the properties whose reality is in question—that is, the properties

represented in color experience. Physicalism must therefore be, in part, a thesis about which properties color experience represents.

### **The Naive Objection**

When the physicalist thesis is so interpreted, however, it tends to elicit the following, naive objection. The microphysical properties of an object are invisible and hence cannot be what is represented when the object looks colored. One can tell an object's color just by looking at it, but one cannot tell anything about its molecular structure-nor, indeed, that it has such a structure-without the aid of instruments or experimentation. How can colors, which are visible, be microphysical properties, which are not?

Physicalists regard this objection as obviously mistaken, although different physicalists regard it as committing different mistakes. A particular physicalist's response to the objection will be conditioned by his brand of physicalism, on the one hand, and his conception of visual representation, on the other. We therefore turn our attention, in the next two sections, to these potential differences among proponents of physicalism.

### **Colors vs. Ways of Being Colored**

The claim that red is a microphysical property can express either of two very different theses. On the one hand, the claim may state a strict identity between properties. In that case, it means that having a particular microphysical configuration is one and the same property as being red. On the other hand, the claim may mean that having this microphysical configuration is a way of being red and, in particular, the way in which things are red in actuality. In that case, the relation drawn between these properties is not identity. Rather, red is envisioned as a higher-order property-the property of having some (lower-order) property satisfying particular conditions-and the microphysical configuration is envisioned as a lower-order property satisfying those conditions, and hence as a realization or embodiment of red.

The difference between these two views is analogous to that between typephysicalism and functionalist materialism in the philosophy of mind. Physicalism says that pain is one and the same state as a configuration of excited neurons. Functionalist materialism says that pain is the higher-order state of occupying some state that plays a particular role, that this role is played in humans by a configuration of excited neurons, and hence that having excited neurons is the way in which humans have pain. Both views can be expressed by the claim that pain is a neural state, but this claim asserts a strict identity only when expressing the former view.

We shall distinguish between the corresponding views of color by referring to them as the physical identity view and the physical realization view, or identityphysicalism and realization-physicalism.<sup>6</sup> To repeat, only identity-physicalism says that red is one and the same property as a

microphysical configuration; realization-physicalism says that the microphysical configuration is merely a way of being red.

Adherents of both identity-physicalism and realization-physicalism will dismiss the naive objection mooted above, but they will dismiss it on different grounds. A realization-physicalist can say that the naive objection confuses color properties with the properties that embody them. The ability to see which color an object instantiates is perfectly compatible, in his view, with an inability to see the particular way in which it instantiates that color. For in his view, seeing that an object is red consists in seeing that it has some property satisfying particular conditions; and seeing that an object has some such property need not entail seeing which such property it has. The invisibility of microphysical properties therefore doesn't preclude them from realizing or embodying colors.

This refutation of the naive objection is not available to the identityphysicalist, of course, since he doesn't draw any distinction between colors and their realizations. The identity-physicalist can still fend off the objection, however, by claiming that it misconstrues the use to which he puts the phrase "microphysical properties." The objection construes this phrase, he says, as articulating a mode of presentation under which colors are represented in visual experience-as expressing what colors are seen as-whereas the phrase is actually intended only to identify the nature of color properties. The physicalist points out that although one never sees anything as a layer of molecules-never sees anything under the characterization "layer of molecules"-one nevertheless sees things that are, in fact, layers of molecules, since that's precisely what the visible surfaces of objects are. Similarly, the physicalist argues, seeing nothing under the mode of presentation "microphysical property" doesn't prevent one from seeing things that are microphysical properties. And that colors are such properties is all that any physicalist means to say.

### **The Propositional Content of Visual Experience**

Thus, the suggestion that physicalism requires colors to be seen under microphysical modes of presentation will be rejected by physicalists of all stripes. Some physicalists will go further, however, by denying that colors are seen under any modes of presentation at all. Whether a physicalist makes this further denial depends on his views about the propositional content of color experience.

On the one hand, a physicalist may take a Fregean view of the visual representation of color. According to that view, the experience of seeing something as red has that content by virtue of the subject's relation to a proposition containing a concept, characterization, or (as we have put it) mode of presentation that is uniquely satisfied by instances of red. The property itself is not an element of the propositional content, as the Fregean conceives it; rather, it is represented by an element of the content, namely, a characterization.

On the other hand, a physicalist might take a completely different view of how color is visually represented, a view that we shall call Russellian. According to that view, the experience of seeing something as red has that content by virtue of the subject's relation to a proposition containing the

property red-the property itself, not a conception, characterization, or presentation of it. A Russellian believes that the property is introduced into the content of experience by something that directly refers to it. This item may be an introspectible, qualitative feature of visual experience, for example, or a word of mentalese tokened in some visual-experience "box." Whatever it is, it must be capable of referring to the property red directly-say, by virtue of a correlation or causal relation with it<sup>7</sup> -rather than by specifying it descriptively, in the sense of having a meaning uniquely satisfied by red objects.<sup>8</sup>

A strict Russellian may believe that the mental symbol for red has no descriptive meaning at all-just a reference. A more liberal Russellian may believe that it has a meaning, but that its meaning is not sufficient to specify the property red or to determine a complete proposition about redness, and hence that the content of seeing something as red must still be completed by the property itself, introduced via direct reference. The difference between these two variants of Russellianism is analogous to that between two variants of the familiar causal theory about natural-kind terms. On the one hand, the word "gold" can be viewed as a name that has no descriptive meaning over and above its reference to gold (although this reference may have been fixed, of course, with the help of a description). On the other hand, "gold" can be viewed as having a descriptive meaning such as "a kind of matter," which is not sufficient to specify a particular kind of matter and must therefore be supplemented by a causally mediated relation of reference to gold. According to the latter view, "gold" and "silver" share the meaning "a kind of matter" but refer to different kinds of matter; and their contributions to the content of sentences must include not only their shared meaning but also their distinct referents. According to the corresponding view about the visual representation of color, there are mental symbols for red and orange that may contribute a shared meaning to the content of visual experiences-say, "a surface property"-while introducing different properties as their referents.

A proponent of this liberal Russellianism will acknowledge that visual experience contains some characterization of colors, but his stricter colleague will not, since the strict Russellian believes that red is introduced into visual content by an item possessing no descriptive meaning at all. The strict Russellian will therefore deny that colors are seen under any modes of presentation. And he will consequently think of the naive objection to physicalism as doubly mistaken-not only in suggesting that he uses the phrase "microphysical property" to articulate such modes of presentation but also in suggesting that he acknowledges their existence.

### **Further Distinctions**

The foregoing responses to the naive objection are cogent, as far as they go; but in our opinion, they don't go far enough. The physicalists have described a way in which microphysically constituted colors aren't represented in visual experience-namely, under microphysical characterizations -but they haven't yet told us how else such colors are represented. Similarly, the realizationphysicalist has described what color properties are not-namely, microphysical properties-but he hasn't yet told us what colors are instead. The realizationphysicalist therefore owes us an account of the higher-order properties that are identical with colors, in his view; and all

of the physicalists owe us an account of how the properties with which they identify colors can be the ones represented in visual experience.

Once again, different physicalists are likely to respond differently. The distinction between Fregeanism and Russellianism and the distinction between the physical identity view and the physical realization view define a four-fold partition of physicalist theories. And within each cell of the partition, further variation is possible. For example, some physicalists believe that the experience of seeing something red normally has a distinctive, introspectible quality in addition to its representational content—a visual "feel," if you will—and that what the experience represents cannot be understood independently of how it feels. Others believe that a visual experience doesn't have intrinsic qualities, or that such qualities are in any case incidental to its content. Different physicalists are also motivated by different intuitions about how physically constituted colors are best identified and hence about how they are likely to be picked out in visual experience. Some identify colors as those physical properties which are common to various classes of objects; they consequently treat the perception of colors as the recognition of physical similarities and differences.<sup>9</sup> Others identify colors as those physical properties which cause particular visual effects, and consequently treat color perception as the recognition, via those effects, of their physical causes.<sup>10</sup>

These disagreements might be thought to require further subdivision of physicalist territory, into eight or even sixteen regions instead of four. But we begin to wonder, at this point, whether all of the resulting regions would be occupied by theories that were even remotely plausible. We shall therefore proceed less abstractly, by developing the latter intuitions about how to identify physically constituted colors. Each of these intuitions could in principle lead to eight different theories, as it is combined with Fregeanism or Russellianism, with identity theory or realization theory, and with credence or skepticism about qualia. As we have suggested, however, not all of the resulting permutations are viable. What's more, the lines of thought departing from these intuitions ultimately tend to converge. We shall therefore attempt to formulate only those accounts of color experience which are both plausible and distinct.

### **The First Intuition: Similarity Classes**

One way of picking out an object as red is by saying that its surface shares a property with the surfaces of ripe tomatoes, British phone booths, McIntosh apples, and so on. Perhaps, then, an object can be visually represented as colored by being represented as sharing a property with certain other objects.

But do references to phone booths and tomatoes crop up in the visual representation of objects as red? Surely, people can see things as red without even having the concept of a tomato or a phone booth. Of course, this particular problem could be circumvented if each person's visual experience were conceived as characterizing red objects in terms of paradigms familiar to that person. But the resulting conception of visual experience would still be wrong, for two reasons.

First, the experience of seeing one thing as red makes no explicit allusion to other instances of the color, familiar or not. No matter how conversant one is with tomatoes, and no matter how centrally tomatoes may have figured in one's acquisition of color concepts, seeing a red fire engine doesn't appear to be an experience about tomatoes. Second, visual experience never represents objects as having their colors necessarily or trivially, whereas it would represent tomatoes (or some other objects) as necessarily and trivially red if it represented things as red by characterizing them as sharing a property with tomatoes (or with those other objects).<sup>1</sup>

The moral of these observations is not that an object's color isn't visually represented as a property shared with other objects; the moral is simply that if it is so represented, the other objects aren't specified individually. The possibility remains that the experience of an object as red represents it as sharing a property with objects in a set that includes tomatoes but which is specified without reference to them or to any other individual members.

Yet how can the appropriate set of objects be specified in the content of visual experience, if not in terms of its members? To suggest that it be specified in terms of a property characteristic of those members would defeat the point of the current intuition. The point of the intuition is that a color can be represented in terms of a set of objects precisely because it's the only property common to all members of the set. Specifying the set in terms of the property characteristic of its members would therefore require an antecedent capacity to represent the color-which would render specification of the set superfluous.

### **A Humean Proposal**

Nevertheless, the intuition that an object's color is seen as a property shared with other objects can be preserved, with the help of a proposal dating back to Hume's Treatise.<sup>2</sup> Imagine that the experience of seeing an object as red has the indexical character "It's one of that kind," wherein the reference of "that kind" is determined by the subject's disposition, at the prompting of the experience, to group the object together with other objects. If the latter objects do constitute a kind, by virtue of possessing some common property, then the experience will have as its content that the former object belongs to that kind and hence that it possesses the characteristic property-a property that could easily be microphysical or realized microphysically.

This account of color experience is of the liberal Russellian variety, since it suggests that visual content characterizes its object as belonging to a kind, but that the kind in question must be specified by direct reference rather than by a more specific characterization. Direct reference is mediated in this case by a correlation between potential classificatory behavior of a subject, on the one hand, and a microphysically constituted kind of object, on the other. As we have seen, the proposal has no Fregean version, because it requires specification of a kind, and no such specification can be found in the introspectible content of color experience.

## **An Information-Theoretic Proposal**

Here is an alternative way of preserving the first intuition. Imagine that a particular mental symbol is regularly tokened in response to visual encounters with objects of a particular kind, whose members belong to it by virtue of possessing some characteristic property. The symbol may then qualify as indicating-and thus, in a sense, as referring to-the kind with which it is correlated., 3 And tokenings of the symbol may consequently introduce its referent into the content of visual experiences, in such a way that objects are represented as members of the kind to which the symbol refers. Such an experience will naturally be described, on the one hand, as registering the similarity of its object to other members of the kind and, on the other, as attributing to its object the property characteristic of the kind. In a sense, then, the object will be seen as having a property by being visually associated with other objects that have it. A microphysical or microphysically realized property may thus be attributed to an object by way of the object's visually detected similarity to other objects.<sup>14</sup>

## **Introducing Qualia**

Now suppose that the mental correlate of a color category were not some item of a subliminal mentalese but, rather, an introspectible sensation or quale. To begin with, this supposition could simply be appended to the foregoing Russellian account. The visual sensation associated with the appearance of a particular object could then be treated like a numeral in a paint-by-numbers scene, assigning the object to a kind, and hence attributing to it an associated property, without characterizing the kind or property in any way. Which kind or property a particular sensation denoted would be fixed, as before, by causation or correlation.<sup>15</sup>

Once the mental correlate of a color category is imagined as accessible to introspection, however, the resources for a Fregean theory become available. The content of a visual experience can then be imagined to invoke the accompanying sensation and hence to characterize its object under the description "having the property that is this sensation's normal or predominant cause."

Such an account of how colors are represented can be adopted by proponents of both identity- and realization-physicalism. An identity -physicalist can say that red is the property referred to within the proposed characterization-the property that tends to cause the accompanying sensation. A realization-physicalist can say that red is the higher-order property expressed by the entire characterization-the property of having a property that tends to cause the sensation. On the first reading, colors may turn out to be identical with microphysical properties; on the second, they may turn out to have microphysical realizations.

## **The Second Intuition: Causes of Visual Effects**

At this point our development of the first intuition, that colors can be identified in terms of similarity classes, has brought us around to the second intuition, that colors can be identified in

terms of their visual effects. Indeed, we have already canvassed the only plausible theories derivable from the latter intuition—namely, theories according to which colors are visually represented by, or by reference to, visual sensations that they cause.

These theories can be paraphrased as saying that colors are visually represented as the properties that normally cause objects to look colored. But such a paraphrase will make sense only if looking colored is understood to consist in giving a visual appearance that's accompanied by particular visual sensations, rather than in being visually represented as having colors. For if colors were represented as the properties that normally cause objects to be represented as having colors, the content of color experience would be viciously circular.

Now, some philosophers have denied that this circularity would be vicious. One philosopher has even claimed that it would be a virtue, in that it would account for the notorious indefinability of colors. Colors are indefinable, he says, precisely because their definitions are unavoidably circular.<sup>16</sup>

We think, however, that the proposed circular definition would imply that the content of color experience is vacuous. When one describes an object as having properties that would cause it to be visually represented as red, one is describing it in terms of the experiences that it is equipped to cause, and one is describing those experiences in terms of their content—namely, as experiences of seeing the thing as red. The content of one's description therefore includes, as a proper part, the content of the experiences that the thing is described as equipped to cause; and the content of one's whole description depends on that component. For this reason, the description cannot express the content of the experiences in question. If the content of seeing something as red were that the thing was equipped to cause experiences of seeing it as red, then the content of seeing something as red would include and depend upon the content of experiences of seeing it as red. The content of seeing something as red would thus include and depend upon itself; it would characterize the thing, in effect, as having a property that would cause experiences containing this very characterization; and hence it would fail to attribute any particular property to the object. Circularity in the content of color experience would render that content vacuous.<sup>17</sup>

Thus, the content of visually representing something as colored cannot be that the thing has whatever normally causes objects to be visually represented as colored. As we have seen, however, the content in question can still be that the thing has whatever normally causes objects to look colored, in the sense that it causes their visual appearances to be accompanied by a color sensation.

### **Outline of the Argument**

We have now developed various proposals for ways in which visual experience might represent microphysically constituted color properties. We began with the Humean proposal that colors are directly denoted by the subject's classificatory dispositions. We then introduced an information-theoretic proposal, which says that colors are directly denoted by mental correlates, whether they be items of mentalese or introspectible qualia. We concluded with a Fregean variant of the latter

possibility, to the effect that colors are characterized descriptively as the properties that normally cause color sensations.

Despite the diversity of these proposals, we think that they are uniformly unsuccessful in showing that visual experience might represent microphysical or microphysically realized colors. Each of them fails to satisfy one of two fundamental requirements for an adequate theory of color vision.

First, we shall argue, a theory of color must respect the epistemology of color experience: it must be compatible with one's knowing what one knows about color properties on the basis of seeing them. The epistemological problem for physicalism is not that the microphysical nature of colors cannot be known on sight; it is rather that other things about colors are known on sight but could not be known in this way if physicalism were true.

Second, we shall argue that a theory of color must respect the phenomenology of color experience: it must be compatible with what it's like to see the world as colored. Mere reflection on what it's like to see colors does not reveal whether the properties being seen are microphysical, but it does yield various constraints on any theory of what those properties are. In particular, such reflection reveals that color experience is naive, in that it purports to acquaint us directly with properties of external objects. In our opinion, no physicalist theory can meet this phenomenological constraint while meeting those imposed by the epistemology of color as well. We consider these constraints in turn, beginning with the epistemological.

## **EPISTEMOLOGICAL CONSTRAINTS**

What do you know about colors, not as a student of physics or physiology, but simply in your capacity as a subject of visual experience? We think that you know, for example, that red and orange are properties; that they are different properties, though of the same kind-different determinants of the same determinable; that they are not as different from one another as they are from blue; and that they cannot simultaneously be instantiated in exactly the same place. Finally, you know that red and orange are properties that things visually appear to have, and you know when things appear to have them.

All but the last two items of knowledge are necessary propositions. Red and orange-that is, the properties that things appear to have in looking red and in looking orange-not only are distinct, similar determinates of the same determinable but are essentially so. A property that wasn't a determinate of the same determinable as red, or wasn't distinct from red, or wasn't similar to red-such a property simply wouldn't be orange. And vice versa.

What's more, mere reflection on color experience provides all the support that might ever be needed for all of the knowledge cited above. That is, you need only reflect on the experiences of seeing things as red and as orange in order to know that they are two distinct, incompatible, but rather similar determinates of a single determinable property; you need only reflect on particular experiences in order to tell which of these properties they represent; and there are no possible circumstances under which more evidence would be needed. We wish to remain neutral on the explanation for this phenomenon. The knowledge in question may be delivered in its entirety by introspection on the contents of the relevant experiences. Alternatively, it may require the recognition of relations among the contents of these experiences, so long as the relations are such as can be recognized a priori. It may even require empirical support, so long as the support required is no more than what's provided by the experiences themselves. All we claim is that the experiences of seeing red and orange provide whatever is necessary for this rudimentary knowledge about those properties.

Consider the consequences of denying that your knowledge about colors has this status. If the experiences of seeing red and orange didn't provide all of the support required for the knowledge that they're distinct but similar determinates of the same determinable, then your knowledge of these matters would be hostage to future empirical discoveries. You would have to consider the possibility of obtaining evidence that red and orange are in fact the same property or, conversely, that they aren't similar at all. And given how the references of "red" and "orange" are fixed, evidence that red and orange are the same property, for example, would amount to evidence that the property that objects appear to have in looking red is the same as the property that they appear to have in looking orange.

Does visual experience leave room for the hypothesis that things appear to have the same property in looking red as they do in looking orange? We think not. Nor does it leave room for the hypothesis that red and orange are less alike than red and blue, or that something seemingly seen as red on a particular occasion is being represented as having a property other than red. Your knowledge on these matters is such that nothing would count as evidence against it.

### **Meeting the Epistemological Constraints**

Yet would such knowledge be possible if physicalism were true? We believe that the answer may be yes in the case of Fregean realization-physicalism, but that in the case of all other versions of physicalism-that is, Russellian theories and identity theories-the answer is no.

What sets the latter theories apart from Fregean realization-physicalism is their implication that visual experiences like yours represent colors only as a matter of contingent fact. Under the terms of these theories, an experience internally indistinguishable from your experience of seeing something as red might fail to represent its object as having that color. The reason is that red is represented by your experience, according to these theories, only by virtue of facts incidental to the internal features of the experience.

Which facts these are depends on the physicalist's conception of visual representation. Under the terms of Russellianism, they are the causal or correlational facts by virtue of which some mental item, or some behavioral disposition, introduces the microphysically constituted property red into the contents of experiences. Twin-earth examples, in the style of Putnam's will readily demonstrate that the same mental item or the same classificatory behavior might have been correlated with objects of a different kind, sharing a different property-in which case, internally similar experiences would not have represented the property that, according to physicalism, is red.

Under the terms of Fregeanism, the facts in virtue of which visual experience represents a microphysical property are the facts in virtue of which instances of that property uniquely satisfy the characterization by which things are visually represented as red. And these facts, too, are bound to be contingent if red is identical with a microphysical property, for reasons illuminated by the naive objection discussed above. Although the naive objection cannot defeat physicalism, it does force the Fregean identity-physicalist to concede that the characterization by which things are visually represented as red does not represent what it is to be red. For as an identity -physicalist, he believes that to be red is to have a particular microphysical property, and yet the objection forces him to concede that things aren't seen under microphysical characterizations. The Fregean identity-physicalist must therefore believe that things are seen as red by means of a contingent characterization-a characterization that is, in fact, uniquely satisfied by instances of the property red, but not because it represents what redness is. And twin-earth examples will once again demonstrate that such a characterization might not have been uniquely satisfied by instances of red or might have been uniquely satisfied by instances of another property. Just as a mental symbol might have tracked a different property, so the visual characterization "whatever causes this feeling" might have been satisfied by a different property; and in either case, your visual experiences wouldn't have represented red, under the terms of the corresponding theory.

Thus, Fregean identity-physicalism is like Russellian physicalism in implying that your experience of something's looking red might have been exactly as it is, in all respects internal to you, while failing to represent anything as red. And this consequence has the corollary that there are circumstances under which you couldn't tell, by mere reflection on the experience of something's looking red, whether it is being represented as having the property red.

The physicalist may object, at this point, that something's being contingent doesn't entail its being a posteriori. He will argue, more specifically, that the reference of "red" has been fixed for you by a description alluding to your visual experiences: red is, by stipulation, whatever property is attributed to objects by their looking red. That red is the property that something appears to have in looking red is therefore knowable a priori, even though it is contingent, just like the length of the standard meter-bar in Paris.<sup>19</sup>

This response misses the epistemological point. The term "red" has been stipulated as denoting the property attributed to objects by their looking red; but the phrase "looks red" has been stipulated as denoting experiences introspectively similar to some paradigm experience. The problem is that under the terms of the theories now in question, there is no introspectively recognizable kind of

experience for which you can always tell by introspection whether the same property is represented in all or most experiences of that kind. These theories therefore imply, to begin with, that, for all you know by reflection on visual experience, the attempt to fix the referent of "red" as the property attributed to objects by their looking red may have failed, since there may be no property that predominantly satisfies that description. They imply furthermore that, even if there is a property represented by most instances of things' looking red, you cannot necessarily tell by reflection when a particular experience is representing that property.

This problem can best be illustrated by imaginary cases of context-switching.<sup>20</sup> Suppose that your environment were to change in such a way that your mental designator for red was correlated with, or your visual characterization of red was satisfied by, a new and different property that replaced the current property red wherever it occurred. At first the content of your visual experiences might remain the same, with the result that you saw objects as having a property that they no longer had. But gradually your visual designators or characterizations would come to denote the new property rather than the old. Tomatoes would therefore appear to have a new and different color property—appear to have it, that is, in the only sense in which a Russellian or an identity-theorist can conceive of them as appearing to have any color at all. Yet in all respects internal to you, your experiences would remain unchanged, and so you would be unable to tell by reflection that you were no longer seeing tomatoes as having the color property that you had previously seen them as having.

Russellianism and identity-physicalism therefore entail that without investigation into the physical causes and correlates of your visual experiences, you cannot necessarily know whether tomatoes appear today to have a different color property from the one that they once appeared to have. You might know that whatever property they appear to have is likely to be the current holder of the title "red," if any property is. But you may not be able to tell when things have appeared to have that property in the past; and you may not be able to tell in the future when things appear to have it. Hence there remains a significant sense in which you don't necessarily know when things appear to be red.<sup>21</sup>

Indeed, these theories entail that you cannot always tell without investigation whether objects appear to have any color properties at all. For just as experiences internally indistinguishable from yours might represent different properties, so too they might simply fail to represent properties. Such a failure would occur if the characterizations applied to objects in visual experience were not satisfied, or if the corresponding mental designators were not systematically correlated with visual stimulation from objects of any particular kind.

Consider the Russellian theories, which say that visual experience represents objects as colored by means of symbols or behavioral dispositions that designate microphysically constituted kinds. Reflection on such an experience wouldn't necessarily reveal whether the symbols being tokened, or the behavior being prompted, were correlated with objects sharing a common property and constituting a genuine kind. For all one could tell from having the experience, the objects associated with the symbol or behavior might be utterly miscellaneous, and so these purported designators

might not indicate membership in a kind or possession of a property. Hence one would be unable to tell, when things looked red, whether there was a property that they thereby appeared to have. And if one didn't know whether things appeared to have a property in looking red, one wouldn't know whether there was such a property as red at all.

The same problem attends the Fregean theory, in all but its realizationtheoretic form. According to Fregean identity-physicalism, as we have developed it, visual experience represents red by characterizing it as the property that normally causes a particular sensation. Yet reflection on a visual representation of this form would not necessarily reveal whether there was a property that normally caused the sensation, and so it wouldn't reveal whether the associated characterization succeeded in denoting a property.

Of course, the possibility of there being no colors represented in visual experience is only the most bizarre of many possibilities that introspection could not rule out if the present theories were true. A less bizarre possibility is that visual experience might represent only two color properties—one when things look either red, orange, or yellow, and another when they look either green, blue, or violet. The correlational or causal facts could certainly be arranged in such a way as to give these experiences one of only two contents, under the terms of Russellian or identity-theoretic physicalism. These theories therefore imply that one cannot always tell without investigation whether red and orange are different colors, the same color, or no color at all.

### **Some Defenses and Replies**

Now, physicalists sometimes admit that visual experience, as they conceive it, is compatible with the possibility that there are no colors.<sup>22</sup> We wonder, however, whether the full import of this concession is generally appreciated. The statement that there may be no colors sounds as if it should gladden the heart of an anti-realist, but it is in fact different from, and perhaps even incompatible with, the views that many anti-realists hold. What these anti-realists believe is that colors are properties that visual experience attributes to objects even though no objects instantiate them. What proponents of the present theories must concede, however, is that there may be no properties attributed to objects by their looking colored, and hence that there may be no such properties as colors, not even uninstantiated ones. They must allow that color experience not only may attribute properties to objects that don't have them, as the anti-realists claim, but may actually fail to attribute properties to objects at all, by failing to express any properties. If this possibility were realized, color experience would lack the representational competence required to be false, strictly speaking, whereas the falsity of color experience is what anti-realism is usually about. And in our opinion, the fact that color experience can at least be false is evident on the face of it.

A physicalist might respond that the designators and characterizations involved in color experience can be assumed to indicate some properties, since something or other is bound to be responsible for one's visual sensations, as specified in the characterizations, and something or other is bound to be correlated with the designators. But the liberal criteria of visual representation that would enable one to assume that some properties or other were being represented would

simultaneously undermine one's claim to other items of knowledge about those properties.<sup>23</sup> For if one's experiences of things as red and as orange represented whatever properties in heaven or earth were correlated with two different designators, or responsible for two different sensations, then one would be even less able to tell by reflection whether those properties belonged to the same determinable, or required extension for their instantiation, or bore greater similarity to one another than to some third property. For all one could tell from seeing colors in the way imagined here, red might be an electrical charge, orange a degree of acidity, and blue a texture.

A physicalist might respond that if the similarities and differences among colors were conceived as relative to an observer, then they would indeed be revealed by reflection on visual experience.<sup>24</sup> Let the imperfect similarity between red and orange consist in the fact that they have distinct but similar effects on normal human observers, and any normal human observer will be able to detect their relation on sight.

The problem with this suggestion is that it can account only for our knowledge of contingent similarities and differences. Red and orange, as conceived by the physicalist, are properties that happen to have distinct but similar effects on human observers, but they might have had effects that were not distinct or were even less similar. Hence the similarity relation that would be accessible by reflection on visual experience, according to physicalism, is a relation that red and orange might not have had. In reality, however, reflection on the experiences of seeing red and orange tells us that if two properties didn't stand in precisely this relation, they wouldn't be the properties we're seeing.<sup>25</sup>

### **Smart's Analogy**

Now, the epistemology of color similarities and differences has received considerable attention from some physicalists who are aware that their theories appear unable to account for it. Because these physicalists subscribe to Russellian or identity-theoretic versions of physicalism, they are committed to the proposition that visual experience doesn't characterize objects in terms that would reveal wherein their color properties consist. The problem is that if visual experience doesn't reveal wherein colors consist, it cannot reveal wherein they are essentially alike or different. In order for visual experience to represent how being red is essentially similar or dissimilar to being orange, it would have to represent what it is to be red or to be orange—which it doesn't do, under the terms of the theories in question. These theories therefore seem unable to explain why the similarities and differences among colors can be known on sight.

Physicalists have attempted to meet this challenge by disputing its premise—namely, that visual experience would have to represent the nature of color properties in order to reveal their similarities and differences. They insist upon "the possibility of being able to report that one thing is like another, without being able to state the respect in which it is like."<sup>26</sup>

J. J. C. Smart once offered an analogy to illustrate this possibility. He wrote:

If we think cybernetically about the nervous system we can envisage it as being able to respond to certain likenesses ... without being able to do more. It would be easier to build a machine which would tell us, say on a punched tape, whether or not ... objects were similar, than it would be to build a machine which would report wherein the similarities consisted.

David Armstrong quotes this passage in application to color similarities and concludes, "No epistemological problem, then."<sup>27</sup>

What Armstrong seems to be suggesting is that one detects the bare fact that red and orange are similar by means of a sensory mechanism that responds to their similarity and produces an awareness of it in one's mind. This similarity-detecting component of the visual sensorium is what corresponds, in Armstrong's view, to the similarity-detecting machine described by Smart.<sup>28</sup> Unfortunately, such a detector, though perfectly conceivable, would not yield the right sort of knowledge about color similarities. For if the similarities among colors were detected by sight, then one's knowledge of them would be defeasible, by evidence of an optical illusion or malfunction. The experience of seeing things as red and as orange would reveal that these colors looked similar, and hence that they were similar if one's eyes could be trusted; but one would have to acknowledge the possibility that their apparent similarity might be an illusion, and that they might not be similar, after all.

In reality, of course, the similarity between red and orange is known beyond question and could not turn out to be an illusion. One needs to have seen red and orange in order to know that they're similar, of course, but only because one needs to have seen them in order to know which properties they are. Once acquainted with them, one doesn't depend on visual evidence for one's knowledge of their similarity, since nothing would count as counter-vailing evidence.

### **Armstrong's Analogy**

Armstrong has suggested that one's ability to perceive color similarities without perceiving their bases is analogous to the ability to perceive family resemblances:

How can we be aware of the resemblance and the incompatibility of the colourshades, yet be unaware of, and have to infer, the nature of the colour-properties from which these features flow? The answer, I take it, is in principle the same ... as for the cases where resemblance of particulars such as faces is observed but the respect of resemblance cannot be made out. Despite the fact that the respect in which the faces resemble one another is not identified, it can still act upon our mind, producing in us an awareness of resemblance.<sup>29</sup>

Now, if we follow Armstrong's instructions to interpret this analogy as comparing the perceived similarity of color properties to the perceived similarity of particular faces, then it does nothing to overcome our stated objection. Although one can often see that two faces are alike, one remains

aware that the appearance of likeness may be illusory, and hence that the faces may turn out not to be alike, after all, whereas the appearance of similarity between red and orange is not subject to empirical refutation.

Yet Armstrong's analogy is open to a slightly different interpretation, which might seem to suggest a case in which knowledge of bare resemblance need not be defeasible, either. Let the similarity between colors be compared not to that between particular faces but, rather, to that between the contours that the faces appear to have, which are properties rather than particulars.<sup>30</sup> When the perception of family resemblance is thus interpreted as the perception of similarity between complex shapes, it no longer seems exposed to the risk of illusion. The faces may not have the shapes that they appear to have, of course, but the similarity between those shapes remains unmistakable, even though one may not be able to articulate the respects in which they're alike. Why, then, can't the similarity between perceived colors be equally unmistakable and yet equally unanalyzable?

The problem with this version of Armstrong's analogy is that one's ignorance of the respects in which perceived shapes are alike is not analogous to the ignorance that one would have of color similarities if Russellian or identitytheoretic physicalism were true. Although one cannot say what's common to the contours that two faces appear to have, one sees those contours under modes of presentation that represent their nature, since shapes are spatial properties and are visually characterized in spatial terms.<sup>31</sup> Information about the aspects in which shapes are similar is therefore included in the introspectible content of their visual appearance. One may just be unable to isolate that information or extract it or put it into words. Under the terms of Russellianism or identity-physicalism, however, one's inability to tell what colors have in common isn't due to the difficulty of processing information contained in their visual characterization; it's due to the absence of that information, since colors aren't characterized in terms that represent their nature.

The difference between these cases is like that between purely referential concepts, which have no sense, and concepts whose sense is difficult to explicate. If one has the concept of gold without being able to say what gold is, the reason may be that having the concept consists in nothing more than standing in the right causal relation to the appropriate objects. But if one has the concept of compassion without being able to say what compassion is, the reason is probably that one's concept has a *de dicto* content that one cannot immediately explicate. Thus, reflection on one's concepts of compassion and pity may not reveal how compassion and pity are alike, any more than reflection on concepts will reveal the relation between gold and silver-but not for the same reason. In the case of gold and silver, the reason will be that one's concepts simply don't reflect the basis of similarity; in the case of compassion and pity, it will be that a relation reflected in one's concepts isn't easy to articulate.

This difference is manifested by differences in one's authority about proposed accounts of the relevant objects or similarities. If someone proposes an account of what gold is, or how it is like silver, one cannot confirm his account simply by consulting one's concepts. But if someone proposes an account of what compassion is, or how it is like pity, reflection on one's concepts may

indeed suffice to reveal whether he's right, even if it wouldn't have enabled one to formulate the account on one's own.

To judge by this test, the visual representation of shape is like a concept that's difficult to explicate, since one can indeed confirm an account of the resemblance between two faces by reflecting on how they look. There is thus good reason to believe that one's knowledge of family resemblance depends on visual information of a sort that is not contained in the appearance of colors, as understood by Russellian or identity-theoretic physicalism. One does see the respects in which two faces are alike, although one may be unable to isolate or describe them, whereas the versions of physicalism under discussion imply that the respects of similarity between colors are utterly invisible. Hence one's ability to be certain about family resemblances is no indication that one could be equally certain about color resemblances if these versions of physicalism were true.

### **Explaining the Epistemological Intuitions Away**

We believe that the foregoing epistemological objections rule out any theory that portrays visual experience as representing colors contingently—that is, without characterizations that denote them necessarily. They thereby rule out Russellian versions of physicalism and Fregean identity-physicalism as well.

Although such theories cannot respect ordinary intuitions about the epistemology of color, some of them can attempt to explain those intuitions away. In particular, any physicalist who acknowledges the existence of distinctive color sensations, or qualia, can argue that we have mistaken introspective knowledge about those sensations for knowledge about the color properties that they help to represent. What the ordinary observer knows by reflection, this physicalist may claim, is not that there are distinct but similar properties that red-looking and orange-looking objects appear to have but, rather, that there are distinct but similar sensations that accompany these appearances. According to this response, we have displaced—indeed, projected—these items of knowledge from their true objects, which are color qualia, onto color properties.

But can the physicalist extend this explanation to our most fundamental knowledge claim, that color experience can be known on reflection to represent properties? He can try. For he can say that we have mistaken the introspectible presence of color qualia in visual experience for an introspectible representation of color properties. Because reflection on visual experience does reveal that things look colored in the sense that their visual appearances are accompanied by color sensations, the physicalist may argue, we have mistaken it as revealing that they look colored in the sense of being represented as having color properties.

But why would we commit this mistake in the case of color, when we have no tendency to commit it in the cases of other, equally vivid sensations? One isn't tempted to think that sensations of pain, for example, attribute any properties to the objects that cause them. Reflection on the experience of being pricked by a pin doesn't yield the conviction that the pin is being represented

as having a pain-property. Why, then, should reflection on an experience accompanied by a color sensation yield the conviction that its object is being represented as colored?

Here again the physicalist may think that he has an explanation. For as Wittgenstein pointed out, sensations of pain, unlike sensations of red, are not regularly received from particular objects or surfaces; if they were, "we should speak of pain-patches just as at present we speak of red patches."<sup>32</sup> Perhaps, then, we believe that visual experience attributes color properties to objects because we've observed the regularity with which color sensations are associated with the perception of particular surfaces. According to this explanation, the knowledge that we have claimed to possess on the basis of mere reflection is in fact derived from observed patterns and correlations within visual experience.<sup>33</sup>

Unfortunately, the patterns and correlations cited here would provide no grounds whatsoever for believing that visual experience attributes color properties to objects in the ways required by Russellian or identity-theoretic physicalism. From the fact that particular objects are individually associated in visual experience with a particular sensation, no conclusion can be drawn about whether the sensation has any normal or predominant cause, and hence about whether there is an external property that it can help to represent. Various objects regularly occasion sensations of red, but those objects are so various that they may not have any surface properties in common, for all one can tell from visual experience. Hence their observed association with one and the same quale provides no grounds for thinking that the quale has any informational potential.

What's more, the association of color sensations with particular objects is no more regular or reliable than that of pain with particular kinds of events. After all, pain serves its monitory function only because young children can learn that it regularly accompanies bumps, scrapes, punctures, encounters with extreme heat or cold, and so forth. Having obvious external correlates is essential to the evolutionary purpose of pain. If what led us to view a sensation as the representation of something external were its observed correlation with various external stimuli, we would have no more occasion to take this view of color than of pain.

Thus, the point of Wittgenstein's remark about pain patches cannot be that pain appears to have no representational content because it has no apparent external correlates. What, then, is the point? Surely, it's that sensations like pain (and color) involve qualities that we can easily think of as located in the external world, but that this thought is blocked, in the case of pain, by there being no particular places where it seems to be located. The external correlates of pain aren't places, and so pain isn't subject to the sort of displacement that the mind practices on other sensations.

Thus, what the association of sensations with particular surfaces produces, and what Wittgenstein was suggesting that it would produce even in the case of pain, is a tendency to perceive the sensations as located on those surfaces—an inducement, in short, to the projective error. But the result of this error is precisely that the qualia themselves, rather than microphysical properties, are attributed to objects in visual experience. Thus, if the patterns cited by the physicalist have their most likely result, they result in the falsity of physicalism as an account of the properties that visual experience represents.

The physicalist explanation of our basic epistemological intuition is therefore unstable. The physicalist wishes to claim that visual experience does not project sensations onto external objects, as their perceived properties, but that reflection on visual experience does project our knowledge about sensations onto objects, as knowledge about their perceived properties. What is cited as accounting for the latter projection doesn't really account for it, however, and would in fact account for the former projection instead.

## **FREGEAN, REALIZATION-THEORETIC THEORIES**

Russellian and identity-theoretic versions of physicalism fail to cope with the epistemology of color because they must portray visual experience as representing color without a characterization that denotes it necessarily. Such visual representations would denote properties only contingently, and would therefore fail to provide the appropriate introspective knowledge of the properties denoted.

This problem does not affect Fregean, realization-theoretic theories. A realization-physicalist can concede, in response to the naive objection, that red objects aren't visually characterized in microphysical terms, and yet hold a version of Fregeanism according to which they are characterized in terms that express what it is to be red; for he doesn't believe that to be red is to have a microphysical property. His theory of visual representation may then enable him to account for the epistemology of color experience. For if visual experience represented red by means of a characterization that represented what it is to be red, then introspection on the content of such an experience would leave no doubt whether there was such a property, introspection on experiences containing the same characterization would leave no doubt whether they represented the same property, and introspection on experiences containing characterizations of various colors would reveal the relations of similarity among them-all because the introspectible content of each experience would reveal what it is to have the property therein represented.

We therefore turn to a consideration of Fregean, realization-theoretic versions of physicalism. One such theory was already introduced, in our initial survey of physicalist theories. Before returning to that theory, however, we shall briefly introduce a new proposal, which is motivated by epistemological arguments of the sort considered above. This proposal has little intuitive appeal of its own; indeed, it would hardly have been intelligible before our epistemological arguments against the other proposals had been aired. As a response to those arguments, however, it has some apparent plausibility.

### **A New Proposal**

The new proposal is an attempt to kill two birds with one stone.<sup>34</sup> It purports to explain at a stroke how colors are visually represented and how their similarities and differences are known. The explanation is that colors are visually characterized precisely as those properties which bear the appropriate similarities and differences to one another.

How could all of the similarities and differences among colors be included in their visual characterization? Here is how.

Let a pigmentation be any property of extended things that stands with its codeterminates in relations of similarity and difference representable by a spheroid space in which distance around the circumference, distance from the ends, and distance from the interior correspond to differences in three different respects (to be called, for our purposes, hue, lightness, and saturation). Then let coordinates be defined so that any determinate pigmentation can be labelled by three numbers specifying its longitude, latitude, and depth in the property space. The description "pigmentation  $xvz$ " will then have as its condition of satisfaction the presence of a determinate whose relation to its co-determinates corresponds to position  $x,vz$  in a property space of this structure.

Now suppose that visual experience characterized surfaces as having pigmentations, specified by their coordinates in pigmentation-space.<sup>35</sup> Under the terms of Fregeanism, such experiences would represent the surfaces as having some appropriately related determinates of some appropriately structured determinable. Under the terms of realization-physicalism, colors would be the second-order properties expressed by such characterizations—that is, the properties of having appropriately related determinates of an appropriately structured determinable.

We believe that this version of the proposal can account for all of the knowledge claimed in our epistemological intuitions. Reflection on the visual characterization of objects as having pigmentations  $xyz$  and  $qrs$  would yield the appropriate knowledge about the higher-order properties that the objects were thereby seen as having. That is, it would reveal that the experience represented its objects as having genuine, co-determinate properties, properties identical to those represented by internally similar experiences and differing from one another in degrees proportionate to  $x - q$ ,  $y - r$ , and  $z - s$ . One would therefore know when one was seeing things as red or orange, and one would be able to tell their similarities and differences.

Unfortunately, this remedy for earlier epistemological problems only creates new ones. Once all of the requisite information has been encoded into the proposed visual characterization of colors, the resulting proposal—in any version—credits the subject of that experience with too much knowledge rather than too little. For it implies that the characterization of any one color encompasses that color's relations to all of the others, by locating it in a fully conceived color space. If color experience conformed to this proposal, the difference between red and orange would not only be evident from the experiences of seeing red and orange; it would be evident from the experience of seeing red alone, since that experience, by representing red as located in a property space of a particular shape, would already intimate the locations of co-determinate properties. The characterization of something as having a property located at longitude  $x$ , latitude  $y$ , and depth  $z$  in a space of co-determinate properties would already suggest the location of properties to the north or south, properties to the east or west, and properties above or below. Yet the experience of seeing something as red does not by itself reveal that the property now in view has a yellower neighbor (orange) and a bluer neighbor (violet), nor that it has more or less bright and more or less saturated neighbors, either. The current proposal has the unfortunate consequence that to see one color is, in

a sense, to see them all.<sup>36</sup> The current proposal thus continues to get the epistemology of color wrong.

### **The Initial Fregean, Realization-Theoretic Proposal**

We therefore return to the initial candidate for a Fregean, realization-theoretic version of physicalism. This was the theory that objects are visually characterized as having properties that normally cause color sensations, and that colors are the higher-order properties expressed by these characterizations. <sup>37</sup>

The content that this theory assigns to visual experience, say, of red would be introspectively recognizable as representing a genuine property; for even if there is no property that's predominantly responsible for sensations of red, the property of having such a property is undoubtedly genuine. Furthermore, any internally similar experience would be introspectively recognizable as representing the same (higher-order) property, by virtue of containing the same characterization. And color properties, so defined, would stand in relations of similarity and difference generated by similarities and differences among the associated sensations. If one visual sensation differed from another in various respects, then the properties of being equipped to cause those sensations would differ isomorphically, by differing as to the sensations caused. Reflection on how it feels to see things as red and as orange would therefore be sufficient to reveal similarities and differences among those colors.

This Fregean, realization-theoretic version of physicalism can therefore account for our knowledge of colors. Unfortunately, it does so at the expense of misrepresenting the phenomenology of color experience.

The present theory implies that the content of visual experience alludes to color qualia as properties distinct from the perceived colors of objects.<sup>38</sup> In order for one to see an object as having the property that causes visual experiences with a particular feel, one's experience would have to represent that feel as well as the property causing it. And one's experience would then lack the naive characteristic of vision.

Visual experience is naive in the sense that it doesn't distinguish between the perceived properties of objects and the properties of perceptions. Whereas the experience of pain, for example, distinguishes between an external cause (a pin's sharpness) and its sensory effect (a finger's pain), visual experience does not distinguish between color as it is in the object and as it feels to the eye: one feels sharp points as causing pains but one doesn't see colored surfaces as causing visual feels. The normal experience of seeing an object as red no more alludes to a sensation as distinct from the object's redness than it does to tomatoes or fire engines.

Thus, the only version of physicalism that gets the epistemology of color experience right gets the phenomenology wrong. In our opinion, any version of physicalism that acknowledges color qualia will commit the same phenomenological error, since it will imply that visual experience always has introspectible color qualities over and above the color properties that it attributes to objects. But this general thesis need not be defended here, since the only version of the second proposal that has survived our epistemological arguments is the Fregean, realization-theoretic version, which portrays visual experience not only as having introspectible qualities but also as alluding to them in its representational content. This version of the proposal implies that visual experience not only involves color sensations but is also about those sensations, in addition to color properties-which is clearly mistaken.

## CONCLUSION

We do not pretend to have proved that any physicalist theory of color must be inadequate, since we have not canvassed every possible theory. We think of our arguments as posing a challenge to any aspiring physicalist. We challenge the physicalist to explain how the physical properties that constitute colors, in his view, are represented in visual experience, and to explain it in a way that meets reasonable epistemological and phenomenological constraints.

One might think that the constraints that we have applied cannot be met by any theory of color, and hence that they must be unreasonable. The solution to the problems we have raised, one might conclude, is not to reject physicalism but rather to relax our epistemological and phenomenological constraints. In our view, however, there is a theory that satisfies these constraints, and it is one of the oldest and most familiar. It is the theory that colors are qualitative properties of visual experiences that are mistakenly projected onto material objects. A defense of that theory must be deferred, however, to another occasion.<sup>39</sup>

<sup>1</sup> Boghossian and Velleman 1989 (this volume, Chapter 13).

<sup>2</sup> Our earlier paper contained a brief discussion of this theory, pp. 82-3 (this volume, pp. 294-5). The present paper can be regarded as expanding on that passage.

<sup>3</sup> These remarks are intended to apply exclusively to expressions of the form “looks colored.” Expressions of the form “seeing something as colored,” “appearing to be colored,” and so forth, will be interpreted compositionally.

Note that the problem discussed in this passage doesn’t preclude “looks red” from *meaning* “visually appears to be red,” in the sense of contributing that content to the statements in which it is used. If the kind of experience denoted by “looks red” is the kind that represents its object as red, then the phrase may indeed be used to introduce the content “visually appears to be red.” The problem discussed here merely restricts the way in which the phrase may acquire its reference to that kind of experience. (See note 17, below.)

<sup>4</sup> Here a further complication arises. We assume in the text that there is a single property represented by all or most instances of looking red. There will certainly be such a property if the way in which experiences qualify as instances of something’s looking red is by representing the same property as the paradigm instance. In that case, the paradigm of something’s looking red will define a kind of experience whose instances attribute the same property to their objects, and so all instances of looking red will represent the same property. The name “red” can then be fixed by the phrase “the property attributed to an object by its looking red.”

But what if kinds of visual appearance are individuated differently? In that case, the kind of appearance defined by a paradigm case of something’s looking red may include appearances representing different properties; and so there may be no property represented in most instances of looking red. Our attempt to attach the name “red” to the property attributed to an object by its looking red will consequently fail, since there will be no single property predominantly satisfying that description.

Physicalists who regard this outcome as a live possibility sometimes think that the reference of “red” cannot be fixed to a single property; and so they make a definition out of the description that we have treated as a reference-fixer. That is, they treat “red” as synonymous with the phrase “the property attributed to an object by its looking red,” and they expect the term, so defined, to denote different properties in different circumstances (see Jackson and Pargetter 1987).

These philosophers may find our usage strange, as we do theirs. But our linguistic differences with them will not prevent us from engaging them in argument. For they believe that red is a microphysical property in some circumstances, in the sense that a microphysical property is the one attributed to an object, in those circumstances, by its looking red. And we shall argue that an object’s looking red never represents it as having a microphysical property, under any circumstances at all. Our arguments will therefore address their view, though not necessarily in their terms.

For a related problem, see the following note.

<sup>5</sup> Some may contend that if there is a physical property that's distinctive of red-looking objects, then it will inevitably be the property that's attributed to objects by their looking red. But this contention simply assumes that the content of color experience is determined in a way that's conducive to the truth of physicalism about color—which should not be assumed from the outset. We shall consider at length whether the content of color experience is determined in this way. At the moment we are merely pointing out that until one has ascertained how colors are visually represented, one must allow for the possibility that their visual representation may have a content that is less useful for people to put into words than other facts correlated with color perception. One must therefore allow for the possibility that the content of color talk may diverge from that of color experience.

Some may argue that even if color experiences somehow represented properties other than external properties correlated with them, they would *also* represent those external properties, by virtue of the correlation. In that case, color experiences would attribute two different properties to their objects, and our identification of colors as the properties attributed to objects by color experiences would be ambiguous. (We owe this suggestion to Sydney Shoemaker.)

What is being imagined here—if it is indeed imaginable—is that visual experiences representing one property would be correlated with another property and would thereby come to indicate it, much as a Cretan's saying "It's raining" may come to indicate sunshine. If visual experiences representing one property could thus come to indicate another, the former property would still be identifiable as the one that they represented *in the first instance* (just as rain would be identifiable as what the Cretan was reporting in the first instance). Our identification of colors as the properties attributed to objects by their looking colored could therefore be easily disambiguated.

<sup>6</sup> Note that in our terminology, identity-physicalism entails that colors have their microphysical natures necessarily. For in our terminology, the thesis "red = microphysical property *x*" is an identity statement whose arguments are rigid designators of properties.

As we explained in note 4, some physicalists treat "red" as synonymous with a non-rigid property description. These physicalists can therefore treat the thesis "red = microphysical property *x*" as a contingent truth.

<sup>7</sup> For correlational theories of reference, see Dretske 1981 and Stampe 1977. We have reservations of a general nature about the prospects for correlational semantics, but we shall suspend these reservations for present purposes. A discussion of this issue can be found in Boghossian 1989b and 1991 (this volume, Chapters 1 and 3 respectively).

<sup>8</sup> For a Russellian view of how colors are visually represented, see Armstrong in Armstrong and Malcolm 1984, p. 172: "A perception of something green will involve a green-sensitive element, that is to say, something which, in a normal environment, is characteristically brought into existence by green things, and which in turn permits the perceiver, if he should so desire, to discriminate by his behaviour the objects from things which are not green."

See also Jackson and Pargetter 1987, pp. 129-30:

What is it for an experience to be the presentation of a property? How must experience E be related to property P, or an instance of P, for E to be the presentation of P, or, equivalently, for E to represent that P? One thing ... is immediately clear. A necessary condition is that there be a causal connection. Sensations of heat are the way heat, that is, molecular kinetic energy, presents itself to us. And this is, in part, a matter of kinetic energy *causing* sensations of heat. We say ‘in part’,

because, for instance, the causation must be in the ‘right way’ For present purposes, however, the causal part of the story is enough. We can work with the rough schema: redness is the property of objects which causes objects to look red

<sup>9</sup> See Smart 1975 and Armstrong 1987.

<sup>10</sup> See Jackson and Pargetter 1987. This view also appears in Shoemaker 1990.

<sup>11</sup> The only way to circumvent this problem would be to suppose that the redness of everything but tomatoes is seen as a surface similarity to tomatoes, whereas the redness of tomatoes is seen as a similarity to fire engines. Yet this supposition would imply that the redness of tomatoes looks different from that of other objects—which is false.

<sup>12</sup> Hume 1978, I.i.vii. This proposal may also be what Armstrong has in mind in some parts of his 1968, chapter 12.

<sup>13</sup> A full correlational theory would identify the referent of a mental item not with its actual causes or correlates but, rather, with the causes or correlates that it would have under counterfactual ideal circumstances. Different theories propose different sets of ideal circumstances, but these differences needn’t concern us here. We shall gloss over these issues by saying simply that under such a theory, a mental item refers to its normal or predominant cause.

<sup>14</sup> This account of how colors are visually represented follows the strict Russellian line, in that it credits the mental symbol for a color with no meaning beyond a correlationally determined reference. There is some room here for liberalization. The mental symbol for red may have a very general sense, such as “a kind of object,” and the visual representation of something as red may therefore characterize it, literally, as of a kind. Which kind is being represented, however, will still be determined by the symbol’s reference, since this account, like the preceding one, offers no resources for a descriptive characterization of the kind.

<sup>15</sup> An identity-theoretic version of this account would say that color sensations denote microphysical properties. A realization-theoretic version would say that they denote higher-order properties that have microphysical realizations.

We do not wish to rule out either of these possibilities entirely. However, one realization-theoretic version of the current proposal can be excluded in advance. This version would be the Russellian counterpart of a Fregean theory that we shall introduce below. The Fregean theory says that visual experience characterizes each color as the higher-order property of having some property that tends to cause a particular color sensation. The Russellian counterpart of this theory would say that each color sensation is appropriately correlated with, and hence refers to, the higher-order property of having a property that tends to cause it.

The problem with the latter theory is that it would utterly trivialize the correlational semantics on which Russellianism depends. Almost every property is correlated with the higher-order property of there being a property that tends to cause it. A semantics that allowed such a correlation to ground a relation of reference would be unable to draw a distinction between what has a reference and what doesn’t.

<sup>16</sup> McDowell 1985.

<sup>17</sup> We develop this argument at length in our 1989, pp. 88-91 (this volume, Chapter 13, pp. 299-302).

Note that the circularity at issue here is significantly different from the circularity at issue in our earlier discussion of the expression “looks red.” There we were concerned with a circularity that could result from the structure of this expression. Identifying red in terms of things’ looking red will be circular, we argued, if “looks red” gets its reference by logical composition, in a way that depends on the reference of “red.” Here we are concerned with a circularity in the content of a visual representation, irrespective of which symbols bear that content or how they are structured. We argue that if the content of representing something as red is that the thing has the property that causes objects to be represented as red, then that content will be embedded in itself. The former circularity can easily be resolved, since “looks red” can be restructured as a unitary expression referring directly to a kind of experience. The latter circularity cannot be resolved by any restructuring of symbols. (See note 3.)

<sup>18</sup> Putnam 1975a; see also Burge 1979.

<sup>19</sup> See Kripke 1972, pp. 54 ff.

Of course, those who define “red” as synonymous with “the property attributed to an object by its looking red” will think that it is not only necessary but analytic that red is the property something appears to have when it looks red. They will therefore claim that their view is compatible with your ability to tell that something appears to be red, since things necessarily appear to be red whenever they look red, and a thing’s looking red is (by stipulation) an introspectively recognizable kind of experience.

True enough. But what these philosophers describe as the ability to tell that something appears to be red is less than meets the ear. It’s the ability to tell that whatever property the thing appears to have is to be called red on this occasion. It’s not the ability to tell when something appears to have that property.

Our claim that you can tell when something appears to be red means that there is a property, red, such that you can tell when something appears to have it. And this claim cannot be accommodated by these linguistic maneuvers.

<sup>20</sup> Context-switching and its relevance to self-knowledge is discussed at greater length in Boghossian 1989a (this volume, Chapter 6).

<sup>21</sup> This problem is especially acute for the Humean proposal, according to which objects are represented as colored by being characterized as “one of *those*,” where the reference of “those” is determined by the subject’s classificatory dispositions. Not only would one be unable to tell by introspection whether the property characteristic of a particular set of objects was the same as it was previously; one would also be unable to tell whether an object was being assigned to the same set of objects as it was previously. In order to tell whether the set to which tomatoes were visually assigned today was the same as the one to which they were assigned yesterday, one would have to investigate precisely which other objects one was disposed to include in that set on each occasion. Mere introspection would therefore fall even further short of revealing whether tomatoes appear to have the same color that they once appeared to have.

<sup>22</sup> See, for example, Armstrong 1968, p. 289.

<sup>23</sup> Such liberal criteria are also unlikely to yield a plausible theory of representation. But as we said in note 7, we are ignoring such general problems in correlational semantics.

<sup>24</sup> We owe this suggestion to Sydney Shoemaker.

<sup>25</sup> Here, as elsewhere, a physicalist may reply that our sense of having introspective knowledge can be explained away. We shall consider this objection below.

<sup>26</sup> These words are from J. J. C. Smart’s “Sensations and Brain Processes.” They are quoted in application to color by Armstrong 1987, p. 12.

<sup>27</sup> Quoted in Armstrong 1987, pp. 12-13.

<sup>28</sup> See Armstrong 1978, p. 127: “[W]hy should not the colour-properties act on our mind (or, rather, why should not states of affairs involving these properties act on our mind),

producing awareness of resemblance and incompatibility, but not producing awareness of those features of the properties from which the resemblance and incompatibility flow?"

Armstrong 1978, p. 127. See also his 1968, pp. 275-6.

29

<sup>30</sup> Armstrong himself sometimes suggests that shapes rather than individuals are the relevant analogue. See Armstrong and Malcolm 1984, pp. 178-9.

<sup>31</sup> One might well have reservations about whether the spatial terms in which shapes are visually characterized fully capture their spatial nature. But such reservations tend to undermine Armstrong's claim that similarities of shape are evident on sight. In assuming that shapes are visually represented in terms that reveal their nature, we are simply taking Armstrong's view of the matter.

<sup>32</sup> *Philosophical Investigations*, p. 312.

<sup>33</sup> This suggestion, too, is due to Sydney Shoemaker.

<sup>34</sup> The following remarks of Armstrong's sound like the theory developed in this section:

The vital point to grasp here, I think, is that, with an exception or two to be noticed, our *concepts* of the individual secondary qualities are quite empty. Consider the colour red. The concept of red does not yield any necessary connection between redness and the surface of ripe Jonathan apples or any other sort of object. It does not yield any necessary connection between redness and any sort of discriminatory behaviour, or capacity for discriminatory behaviour, in us or in other creatures. It does not yield any necessary connection between redness and the way that the presence of redness is detected (eyes, etc.) in us or in other creatures. Finally, and most importantly, it does not yield any necessary connection between red objects and any sort of perceptual experience, such as looking red to normal perceivers in normal viewing conditions.

There may be a conceptual connection between redness and extendedness. There is certainly

a conceptual connection between redness and the other colours: the complex resemblances and differences that the colours have to each other. But these conceptual connections do not enable us to break out of the circle of the colours. (Armstrong 1987, p. 11)

However, the rest of Armstrong's work makes clear that he does not subscribe to the theory developed here. In his 1978, he attributes such a theory to R. W. Church (pp. 108-11).

<sup>35</sup> The use of numerical coordinates is not essential to this conception of color experience. Visual experience can be conceived as locating colors in the property space directly, without the use of coordinates; or it can be conceived as locating them in a network of similarity relations, without the use of any spatial analogy at all.

<sup>36</sup> A proponent of this theory may reply that the complete conception of color space may be acquired gradually, as the subject of visual experience encounters new colors. But this reply isn't to the point, because our objection is not especially about the acquisition of color concepts. What refutes the present theory of color representation is not just that someone who has never seen orange cannot derive the concept of it from seeing red. It's also that someone who has seen both red and orange still does not have experiences of either color that, by themselves, would ground knowledge about the other.

What's more, the most plausible account of how a naive subject might discover color space is not compatible with the proposal under consideration. For according to the proposal, either one already sees colors in terms of their locations in a space of co-determinates—in which case the appearance of one color already alludes to the others—or one doesn't yet see colors in terms of their locations in color space—in which case, their appearances furnish no grounds for drawing the similarities and differences constitutive of such a space. One colored surface will appear to differ from another along three dimensions, according to the proposal, only if each surface is already seen under characterizations specifying three coordinates for its pigmentation. Hence the proposal doesn't allow for the possibility of *discovering* the dimensions of color space *on the basis of* what is seen in color experience. If one doesn't already see colors under characterizations locating them in such a space, then one sees nothing on the basis of which locations could be assigned to them.

<sup>37</sup> Whether one regards realization-physicalism as equivalent to dispositionalism will depend on one's views on the relation between dispositions and their bases. For a dispositionalist theory of color that may be equivalent to realization-physicalism, see Peacocke 1984.

<sup>38</sup> The following argument is developed more fully in Boghossian and Velleman 1989, pp. 94-6.

<sup>39</sup> For comments on earlier drafts of this paper, we are grateful to David Armstrong, C. L. Hardin, David Hills, Sydney Shoemaker, and Steve Yablo.